

## INTERNATIONAL SEARCH REPORT

International Application No.

PCT/ 00/02433

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/12 C12N5/10 C07K14/47 C07K16/18 C12Q1/68  
G01N33/50 A01K67/027

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K A01K G01N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EMBL, STRAND, BIOSIS

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KIKUNO R. ET AL.: "Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 cDNA clones from brain which code for large proteins in vitro." DNA RES., vol. 6, 30 June 1999 (1999-06-30), pages 197-205, XP000852618 the whole document -& DATABASE EMBL 'Online! Accession No. AB029006, 4 August 1999 (1999-08-04) OHARA O.: "Homo sapiens mRNA for KIAA1083 protein" XP002140994 the whole document --- -/--	1,4,5, 14,16, 18-21, 24-30



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

## \* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*G\* document member of the same patent family

Date of the actual completion of the international search

4 January 2001

Date of mailing of the international search report

25/01/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Gal11, 1



## INTERNATIONAL SEARCH REPORT

International Application No

PCT/FR 00/02433

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL SEQUENCES 'Online! Accession No. HS1179930, 29 March 1997 (1997-03-29) MYERS R.M.: "human STS SHGC-44567" XP002156510 brin complémentaire aux nucléotides 2204-2575 de la séquence no. 2 ---	1,4,5, 14,16, 18-21, 24-30
P,X	HAZAN J. ET AL.: "Spastin, a new AAA protein, is altered in the most frequent form of autosomal dominant spastic paraplegia" NAT. GENET., vol. 23, November 1999 (1999-11), pages 296-303, XP000914979 the whole document ---	1-36
A	HEINZLEF O. ET AL.: "Mapping of a complicated familial spastic paraplegia to locus SPG4 on chromosome 2p" J. MED. GENET., vol. 35, no. 2, February 1998 (1998-02), pages 89-93, XP000914971 cited in the application the whole document -----	1-36



Small, faint, illegible marks or characters at the top right of the page.

## Continuation of Box I.2

Claims 13-15, 19-36 partly concern nucleic acid sequences promoting and/or regulating the expression of the SPG4 gene, as well as the corresponding polypeptides, without providing therein a real technical characterisation. Said claims concern also nucleic acids and polypeptides derived from any other gene controlling SPG4, which is neither defined nor characterised in the description. Consequently, said claims are ambiguous and vague, and their subject matter is not sufficiently disclosed and supported in accordance with PCT Article 5 and 6. Consequently, it is not possible to carry out any meaningful search for speculative claims whereof the subject matter consists merely of a list of results to be obtained in future.

The applicant's attention is drawn to the fact that claims, or parts of claims, in respect of which no search report has been established need not be the subject of a preliminary examination report (PCT Rule 66.1(e)). The applicant is advised that the line of conduct adopted by the EPO acting in its capacity as International Preliminary Examination Authority is not to proceed with an examination of a subject matter in respect of which no search has been carried out. Such an attitude will remain unchanged, notwithstanding whether the claims have been modified or not, either after the report has been received or during any procedure under Chapter II.



## TRAITE DE COOPERATION EN MATIERE DE BREVETS

PCT

Expéditeur: le BUREAU INTERNATIONAL

NOTIFICATION DE L'ENREGISTREMENT  
D'UN CHANGEMENT(règle 92bis.1 et  
instruction administrative 422 du PCT)

Destinataire:

MARTIN, Jean-Jacques  
Cabinet Regimbeau  
20 rue de Chazelles  
F-75847 Paris  
FRANCE

Date d'expédition (jour/mois/année) 26 avril 2001 (26.04.01)	NOTIFICATION IMPORTANTE
Référence du dossier du déposant ou du mandataire 341225/18374	
Demande internationale no PCT/FR00/02433	Date du dépôt international (jour/mois/année) 04 septembre 2000 (04.09.00)

## 1. Les renseignements suivants étaient enregistrés en ce qui concerne:

☐ le déposant      ☐ l'inventeur      ☒ le mandataire      ☐ le représentant commun

Nom et adresse MARTIN, Jean-Jacques Cabinet Regimbeau 26, avenue Kléber F-75116 Paris FRANCE	Nationalité (nom de l'Etat)	Domicile (nom de l'Etat)
	no de téléphone 01-45-00-92-02	
	no de télécopieur 01-45-00-46-12	
	no de téléimprimeur	

## 2. Le Bureau international notifie au déposant que le changement indiqué ci-après a été enregistré en ce qui concerne:

☐ la personne      ☐ le nom      ☒ l'adresse      ☐ la nationalité      ☐ le domicile

Nom et adresse MARTIN, Jean-Jacques Cabinet Regimbeau 20 rue de Chazelles F-75847 Paris FRANCE	Nationalité (nom de l'Etat)	Domicile (nom de l'Etat)
	no de téléphone 01-44-29-35-00	
	no de télécopieur 01-44-29-35-99	
	no de téléimprimeur	

## 3. Observations complémentaires, le cas échéant:

## 4. Une copie de cette notification a été envoyée:

☒ à l'office récepteur      ☒ aux offices désignés concernés  
☐ à l'administration chargée de la recherche internationale      ☐ aux offices élus concernés  
☐ à l'administration chargée de l'examen préliminaire international      ☐ autre destinataire:

Bureau international de l'OMPI 34, chemin des Colombettes 1211 Genève 20, Suisse no de télécopieur (41-22) 740.14.35	Fonctionnaire autorisé: Jocelyne Rey-Millet no de téléphone (41-22) 338.83.38
---	---



# PCT

## REQUÊTE

Le soussigné requiert que la présente demande internationale soit traitée conformément au Traité de coopération en matière de brevets.

Réservé à l'office récepteur

Demande internationale n°

Date du dépôt international

Nom de l'office récepteur et "Demande internationale PCT"

Référence du dossier du déposant ou du mandataire (facultatif)  
(12 caractères au maximum) 341225/18374

Cadre n° I **TITRE DE L'INVENTION** **CLONAGE, EXPRESSION ET CARACTERISATION DU GENE SPG4 RESPONSABLE DE LA FORME LA PLUS FREQUENTE DE PARAPLEGIE SPASTIQUE AUTOSOMIQUE DOMINANTE**

Cadre n° II **DÉPOSANT**

Nom et adresse : (Nom de famille suivi du prénom; pour une personne morale, désignation officielle complète. L'adresse doit comprendre le code postal et le nom du pays. Le pays de l'adresse indiquée dans ce cadre est l'État où le déposant a son domicile si aucun domicile n'est indiqué ci-dessous.)

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS)  
3 Rue Michel-Ange  
75794 PARIS CEDEX 16  
FRANCE

☐ Cette personne est aussi inventeur.

n° de téléphone

n° de télécopieur

n° de téléimprimeur

Nationalité (nom de l'État) :  
FR

Domicile (nom de l'État) :  
FR

Cette personne est déposant pour :

☐ tous les États désignés

☒ tous les États désignés sauf les États-Unis d'Amérique

☐ les États-Unis d'Amérique seulement

☐ les États indiqués dans le cadre supplémentaire

Cadre n° III **AUTRE(S) DÉPOSANT(S) OU (AUTRE(S)) INVENTEUR(S)**

Nom et adresse : (Nom de famille suivi du prénom; pour une personne morale, désignation officielle complète. L'adresse doit comprendre le code postal et le nom du pays. Le pays de l'adresse indiquée dans ce cadre est l'État où le déposant a son domicile si aucun domicile n'est indiqué ci-dessous.)

WEISSENBACH Jean  
163 Rue de Vaugirard  
75015 PARIS  
FRANCE

Cette personne est :

☐ déposant seulement

☒ déposant et inventeur

☐ inventeur seulement  
(Si cette case est cochée, ne pas remplir la suite.)

Nationalité (nom de l'État) :  
FR

Domicile (nom de l'État) :  
FR

Cette personne est déposant pour :

☐ tous les États désignés

☐ tous les États désignés sauf les États-Unis d'Amérique

☒ les États-Unis d'Amérique seulement

☐ les États indiqués dans le cadre supplémentaire

☒ D'autres déposants ou inventeurs sont indiqués sur une feuille annexe.

Cadre n° IV **MANDATAIRE OU REPRÉSENTANT COMMUN; OU ADRESSE POUR LA CORRESPONDANCE**

La personne dont l'identité est donnée ci-dessous est : a été désignée pour agir au nom du ou des déposants auprès des autorités internationales compétentes, comme,

☒

mandataire

☐

représentant commun

Nom et adresse : (Nom de famille suivi du prénom; pour une personne morale, désignation officielle complète. L'adresse doit comprendre le code postal et le nom du pays.)

MARTIN Jean-Jacques, SCHRIMPF Robert, AHNER Francis  
WARCOIN Jacques, TEXIER Christian,  
CABINET REGIMBEAU  
26 Avenue Kléber  
75116 PARIS  
FRANCE

n° de téléphone

01 45 00 92 02

n° de télécopieur

01 45 00 46 12

n° de téléimprimeur

☐ Adresse pour la correspondance : cocher cette case lorsque aucun mandataire ni représentant commun n'est/n'a été désigné et que l'espace ci-dessus est utilisé pour indiquer une adresse spéciale à laquelle la correspondance doit être envoyée.



Suite du cadre n° III AUTRE(S) DÉPOSANT(S) OU (AUTRE(S)) INVENTEUR(S)	
<i>Si aucun des sous-cadres suivants n'est utilisé, cette feuille ne doit pas être incluse dans la requête.</i>	
<p>Nom et adresse : (Nom de famille suivi du prénom; pour une personne morale, désignation officielle complète. L'adresse doit comprendre le code postal et le nom du pays. Le pays de l'adresse indiquée dans ce cadre est l'État où le déposant a son domicile si aucun domicile n'est indiqué ci-dessous.)</p> <p>HAZAN Jamilé 22 Rue des Cordelières 75013 PARIS FRANCE</p>	<p>Cette personne est :</p> <p><input type="checkbox"/> déposant seulement</p> <p><input checked="" type="checkbox"/> déposant et inventeur</p> <p><input type="checkbox"/> inventeur seulement (Si cette case est cochée, ne pas remplir la suite.)</p>
Nationalité (nom de l'État) : FR	Domicile (nom de l'État) : FR
<p>Cette personne est déposant pour : <input type="checkbox"/> tous les États désignés <input type="checkbox"/> tous les États désignés sauf les États-Unis d'Amérique <input checked="" type="checkbox"/> les États-Unis d'Amérique seulement <input type="checkbox"/> les États indiqués dans le cadre supplémentaire</p>	
<p>Nom et adresse : (Nom de famille suivi du prénom; pour une personne morale, désignation officielle complète. L'adresse doit comprendre le code postal et le nom du pays. Le pays de l'adresse indiquée dans ce cadre est l'État où le déposant a son domicile si aucun domicile n'est indiqué ci-dessous.)</p>	<p>Cette personne est :</p> <p><input type="checkbox"/> déposant seulement</p> <p><input type="checkbox"/> déposant et inventeur</p> <p><input type="checkbox"/> inventeur seulement (Si cette case est cochée, ne pas remplir la suite.)</p>
Nationalité (nom de l'État) :	Domicile (nom de l'État) :
<p>Cette personne est déposant pour : <input type="checkbox"/> tous les États désignés <input type="checkbox"/> tous les États désignés sauf les États-Unis d'Amérique <input type="checkbox"/> les États-Unis d'Amérique seulement <input type="checkbox"/> les États indiqués dans le cadre supplémentaire</p>	
<p>Nom et adresse : (Nom de famille suivi du prénom; pour une personne morale, désignation officielle complète. L'adresse doit comprendre le code postal et le nom du pays. Le pays de l'adresse indiquée dans ce cadre est l'État où le déposant a son domicile si aucun domicile n'est indiqué ci-dessous.)</p>	<p>Cette personne est :</p> <p><input type="checkbox"/> déposant seulement</p> <p><input type="checkbox"/> déposant et inventeur</p> <p><input type="checkbox"/> inventeur seulement (Si cette case est cochée, ne pas remplir la suite.)</p>
Nationalité (nom de l'État) :	Domicile (nom de l'État) :
<p>Cette personne est déposant pour : <input type="checkbox"/> tous les États désignés <input type="checkbox"/> tous les États désignés sauf les États-Unis d'Amérique <input type="checkbox"/> les États-Unis d'Amérique seulement <input type="checkbox"/> les États indiqués dans le cadre supplémentaire</p>	
<p>Nom et adresse : (Nom de famille suivi du prénom; pour une personne morale, désignation officielle complète. L'adresse doit comprendre le code postal et le nom du pays. Le pays de l'adresse indiquée dans ce cadre est l'État où le déposant a son domicile si aucun domicile n'est indiqué ci-dessous.)</p>	<p>Cette personne est :</p> <p><input type="checkbox"/> déposant seulement</p> <p><input type="checkbox"/> déposant et inventeur</p> <p><input type="checkbox"/> inventeur seulement (Si cette case est cochée, ne pas remplir la suite.)</p>
Nationalité (nom de l'État) :	Domicile (nom de l'État) :
<p>Cette personne est déposant pour : <input type="checkbox"/> tous les États désignés <input type="checkbox"/> tous les États désignés sauf les États-Unis d'Amérique <input type="checkbox"/> les États-Unis d'Amérique seulement <input type="checkbox"/> les États indiqués dans le cadre supplémentaire</p>	
<p><input type="checkbox"/> D'autres déposants ou inventeurs sont indiqués sur une autre feuille annexe.</p>	



## Cadre n° V DÉSIGNATION D'ÉTATS

Les désignations suivantes sont faites conformément à la règle 4.9.a) (cocher les cases appropriées; une au moins doit l'être) :

## Brevet régional

- ☐ AP Brevet ARIPO : GH Ghana, GM Gambie, KE Kenya, LS Lesotho, MW Malawi, SD Soudan, SL Sierra Leone, SZ Swaziland, TZ République-Unie de Tanzanie, UG Ouganda, ZW Zimbabwe et tout autre État qui est un État contractant du Protocole de Harare et du PCT
- ☐ EA Brevet eurasiatique : AM Arménie, AZ Azerbaïdjan, BY Bélarus, KG Kirghizistan, KZ Kazakhstan, MD République de Moldova, RU Fédération de Russie, TJ Tadjikistan, TM Turkménistan et tout autre État qui est un État contractant de la Convention sur le brevet eurasiatique et du PCT
- ☒ EP Brevet européen : AT Autriche, BE Belgique, CH et LI Suisse et Liechtenstein, CY Chypre, DE Allemagne, DK Danemark, ES Espagne, FI Finlande, FR France, GB Royaume-Uni, GR Grèce, IE Irlande, IT Italie, LU Luxembourg, MC Monaco, NL Pays-Bas, PT Portugal, SE Suède et tout autre État qui est un État contractant de la Convention sur le brevet européen et du PCT
- ☐ OA Brevet OAPI : BF Burkina Faso, BJ Bénin, CF République centrafricaine, CG Congo, CI Côte d'Ivoire, CM Cameroun, GA Gabon, GN Guinée, GW Guinée-Bissau, ML Mali, MR Mauritanie, NE Niger, SN Sénégal, TD Tchad, TG Togo et tout autre État qui est un État membre de l'OAPI et un État contractant du PCT (si une autre forme de protection ou de traitement est souhaitée, le préciser sur la ligne pointillée) . . . . .

## Brevet national (si une autre forme de protection ou de traitement est souhaitée, le préciser sur la ligne pointillée) :

- |  |   |
|--|---|
| <input type="checkbox"/> AE Émirats arabes unis                        | <input type="checkbox"/> LR Libéria                               |
| <input type="checkbox"/> AL Albanie                                    | <input type="checkbox"/> LS Lesotho                               |
| <input type="checkbox"/> AM Arménie                                    | <input type="checkbox"/> LT Lituanie                              |
| <input type="checkbox"/> AT Autriche                                   | <input type="checkbox"/> LU Luxembourg                            |
| <input type="checkbox"/> AU Australie                                  | <input type="checkbox"/> LV Lettonie                              |
| <input type="checkbox"/> AZ Azerbaïdjan                                | <input type="checkbox"/> MA Maroc                                 |
| <input type="checkbox"/> BA Bosnie-Herzégovine                         | <input type="checkbox"/> MD République de Moldova                 |
| <input type="checkbox"/> BB Barbade                                    | <input type="checkbox"/> MG Madagascar                            |
| <input type="checkbox"/> BG Bulgarie                                   | <input type="checkbox"/> MK Ex-République yougoslave de Macédoine |
| <input type="checkbox"/> BR Brésil                                     | <input type="checkbox"/> MN Mongolie                              |
| <input type="checkbox"/> BY Bélarus                                    | <input type="checkbox"/> MW Malawi                                |
| <input checked="" type="checkbox"/> CA Canada                          | <input type="checkbox"/> MX Mexique                               |
| <input type="checkbox"/> CH et LI Suisse et Liechtenstein              | <input type="checkbox"/> NO Norvège                               |
| <input type="checkbox"/> CN Chine                                      | <input type="checkbox"/> NZ Nouvelle-Zélande                      |
| <input type="checkbox"/> CR Costa Rica                                 | <input type="checkbox"/> PL Pologne                               |
| <input type="checkbox"/> CU Cuba                                       | <input type="checkbox"/> PT Portugal                              |
| <input type="checkbox"/> CZ République tchèque                         | <input type="checkbox"/> RO Roumanie                              |
| <input type="checkbox"/> DE Allemagne                                  | <input type="checkbox"/> RU Fédération de Russie                  |
| <input type="checkbox"/> DK Danemark                                   | <input type="checkbox"/> SD Soudan                                |
| <input type="checkbox"/> DM Dominique                                  | <input type="checkbox"/> SE Suède                                 |
| <input type="checkbox"/> EE Estonie                                    | <input type="checkbox"/> SG Singapour                             |
| <input type="checkbox"/> ES Espagne                                    | <input type="checkbox"/> SI Slovénie                              |
| <input type="checkbox"/> FI Finlande                                   | <input type="checkbox"/> SK Slovaquie                             |
| <input type="checkbox"/> GB Royaume-Uni                                | <input type="checkbox"/> SL Sierra Leone                          |
| <input type="checkbox"/> GD Grenade                                    | <input type="checkbox"/> TJ Tadjikistan                           |
| <input type="checkbox"/> GE Géorgie                                    | <input type="checkbox"/> TM Turkménistan                          |
| <input type="checkbox"/> GH Ghana                                      | <input type="checkbox"/> TR Turquie                               |
| <input type="checkbox"/> GM Gambie                                     | <input type="checkbox"/> TT Trinité-et-Tobago                     |
| <input type="checkbox"/> HR Croatie                                    | <input type="checkbox"/> TZ République-Unie de Tanzanie           |
| <input type="checkbox"/> HU Hongrie                                    | <input type="checkbox"/> UA Ukraine                               |
| <input type="checkbox"/> ID Indonésie                                  | <input type="checkbox"/> UG Ouganda                               |
| <input type="checkbox"/> IL Israël                                     | <input checked="" type="checkbox"/> US États-Unis d'Amérique      |
| <input type="checkbox"/> IN Inde                                       | <input type="checkbox"/> UZ Ouzbékistan                           |
| <input type="checkbox"/> IS Islande                                    | <input type="checkbox"/> VN Viet Nam                              |
| <input checked="" type="checkbox"/> JP Japon                           | <input type="checkbox"/> YU Yougoslavie                           |
| <input type="checkbox"/> KE Kenya                                      | <input type="checkbox"/> ZA Afrique du Sud                        |
| <input type="checkbox"/> KG Kirghizistan                               | <input type="checkbox"/> ZW Zimbabwe                              |
| <input type="checkbox"/> KP République populaire démocratique de Corée |   |
| <input type="checkbox"/> KR République de Corée                        |   |
| <input type="checkbox"/> KZ Kazakhstan                                 |   |
| <input type="checkbox"/> LC Sainte-Lucie                               |   |
| <input type="checkbox"/> LK Sri Lanka                                  |   |

Cases réservées pour la désignation d'États qui sont devenus parties au PCT après la publication de la présente feuille :

- ☐ DZ Algérie ☐ MZ Mozambique
- ☐ AG Antigua et Barbuda ☐ BZ Belize

**Déclaration concernant les désignations de précaution :** outre les désignations faites ci-dessus, le déposant fait aussi conformément à la règle 4.9.b) toutes les désignations qui seraient autorisées en vertu du PCT, à l'exception de toute désignation indiquée dans le cadre supplémentaire comme étant exclue de la portée de cette déclaration. Le déposant déclare que ces désignations additionnelles sont faites sous réserve de confirmation et que toute désignation qui n'est pas confirmée avant l'expiration d'un délai de 15 mois à compter de la date de priorité doit être considérée comme retirée par le déposant à l'expiration de ce délai. (La confirmation (y compris les taxes) doit parvenir à l'office récepteur dans le délai de 15 mois.)



Cadre n° VI REVENDEICATION DE PRIORITÉ		<input type="checkbox"/> D'autres revendications de priorité sont indiquées dans le cadre supplémentaire.		
Date de dépôt de la demande antérieure (jour/mois/année)	Numéro de la demande antérieure	Lorsque la demande antérieure est une :		
		demande nationale : pays	demande régionale : * office régional	demande internationale : office récepteur
(1) 03/09/99	99 11097	FRANCE		
(2)				
(3)				

☐ L'office récepteur est prié de préparer et de transmettre au Bureau international une copie certifiée conforme de la ou des demandes antérieures (seulement si la demande antérieure a été déposée auprès de l'office qui, aux fins de la présente demande internationale, est l'office récepteur) indiquées ci-dessus au(x) point(s) : \_\_\_\_\_

\* Si la demande antérieure est une demande ARIPO, il est obligatoire d'indiquer dans le cadre supplémentaire au moins un pays partie à la Convention de Paris pour la protection de la propriété industrielle pour lequel cette demande antérieure a été déposée (règle 4.10.b)ii). Voir le cadre supplémentaire.

Cadre n° VII ADMINISTRATION CHARGÉE DE LA RECHERCHE INTERNATIONALE			
Choix de l'administration chargée de la recherche internationale (ISA) (si plusieurs administrations chargées de la recherche internationale sont compétentes pour procéder à la recherche internationale, indiquer l'administration choisie: le code à deux lettres peut être utilisé). <b>ISA/ EP</b>	Demande d'utilisation des résultats d'une recherche antérieure: mention de cette recherche (si une recherche antérieure a été effectuée par l'administration chargée de la recherche internationale ou demandée à cette dernière): Date (jour/mois/année)      Numéro      Pays (ou office régional) 26 JUIN 2000      FA 583429      OEB		

Cadre n° VIII BORDEREAU: LANGUE DE DÉPÔT	
La présente demande internationale contient le nombre de feuilles suivant :  requête : 4 description (sauf partie réservée au listage des séquences) : 44 revendications : 4 abrégé : 1 dessins : 7 partie de la description réservée au listage des séquences : 57  Nombre total de feuilles : 117	Le ou les éléments cochés ci-après sont joints à la présente demande internationale : 1. <input type="checkbox"/> feuille de calcul des taxes 2. <input type="checkbox"/> pouvoir distinct signé <b>à suivre (2)</b> 3. <input type="checkbox"/> copie du pouvoir général; numéro de référence, le cas échéant : 4. <input type="checkbox"/> explication de l'absence d'une signature 5. <input checked="" type="checkbox"/> document(s) de priorité indiqué(s) dans le cadre n° VI au(x) point(s) : 6. <input type="checkbox"/> traduction de la demande internationale en (langue) : 7. <input type="checkbox"/> indications séparées concernant des micro-organismes ou autre matériel biologique déposés 8. <input checked="" type="checkbox"/> listage des séquences de nucléotides ou d'acides aminés sous forme déchiffrable par ordinateur 9. <input checked="" type="checkbox"/> autres éléments (préciser) : Copie du Rapport de Recherche
Figure des dessins qui doit accompagner l'abrégé :	Langue de dépôt de la demande internationale : Français

Cadre n° IX SIGNATURE DU DÉPOSANT OU DU MANDATAIRE	
À côté de chaque signature, indiquer le nom du signataire et, si cela n'apparaît pas clairement, le nom et le titre de l'intéressé signifié.  <div style="display: flex; justify-content: space-between; align-items: flex-start;"> <div style="width: 40%;">   <b>WARCOIN Jacques</b> </div> <div style="width: 50%; text-align: center;">   <b>CABINET RECIMBEAU</b>  <b>CONSEILS EN PROPRIÉTÉ INDUSTRIELLE</b>  <b>28, Avenue Kléber</b>  <b>75116 PARIS FRANCE</b> </div> </div>	

Réservé à l'office récepteur	
1. Date effective de réception des pièces supposées constituer la demande internationale :  3. Date effective de réception, rectifiée en raison de la réception ultérieure, mais dans les délais, de documents ou de dessins complétant ce qui est supposé constituer la demande internationale :  4. Date de réception, dans les délais, des corrections demandées selon l'article 11.2) du PCT :	2. Dessins : <input type="checkbox"/> reçus :  <input type="checkbox"/> non reçus :
5. Administration chargée de la recherche internationale (si plusieurs sont compétentes) : <b>ISA/</b>	6. <input type="checkbox"/> Transmission de la copie de recherche différée jusqu'au paiement de la taxe de recherche.

Date de réception de l'exemplaire original par le Bureau international :	Réservé au Bureau international
--	---------------------------------



# TRAITE DE COOPERATION EN MATIERE DE BREVETS

**PCT**

## AVIS INFORMANT LE DEPOSANT DE LA COMMUNICATION DE LA DEMANDE INTERNATIONALE AUX OFFICES DESIGNES

(règle 47.1.c), première phrase, du PCT)

Expéditeur: le BUREAU INTERNATIONAL

Destinataire:

MARTIN, Jean-Jacques  
Cabinet Regimbeau  
26, avenue Kléber  
F-75116 Paris  
FRANCE



Date d'expédition (jour/mois/année) 15 mars 2001 (15.03.01)		
Référence du dossier du déposant ou du mandataire 341225/18374		AVIS IMPORTANT
Demande internationale no PCT/FR00/02433	Date du dépôt international (jour/mois/année) 04 septembre 2000 (04.09.00)	Date de priorité (jour/mois/année) 03 septembre 1999 (03.09.99)
Déposant CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) etc		

1. Il est notifié par la présente qu'à la date indiquée ci-dessus comme date d'expédition de cet avis, le Bureau international a communiqué, comme le prévoit l'article 20, la demande internationale aux offices désignés suivants:

US

Conformément à la règle 47.1.c), troisième phrase, ces offices acceptent le présent avis comme preuve déterminante du fait que la communication de la demande internationale a bien eu lieu à la date d'expédition indiquée plus haut, et le déposant n'est pas tenu de remettre de copie de la demande internationale à l'office ou aux offices désignés.

2. Les offices désignés suivants ont renoncé à l'exigence selon laquelle cette communication doit être effectuée à cette date:

CA,EP,JP

La communication sera effectuée seulement sur demande de ces offices. De plus, le déposant n'est pas tenu de remettre de copie de la demande internationale aux offices en question (règle 49.1)a-bis)).

3. Le présent avis est accompagné d'une copie de la demande internationale publiée par le Bureau international le 15 mars 2001 (15.03.01) sous le numéro WO 01/18198

### RAPPEL CONCERNANT LE CHAPITRE II (article 31.2)a) et règle 54.2)

Si le déposant souhaite reporter l'ouverture de la phase nationale jusqu'à 30 mois (ou plus pour ce qui concerne certains offices) à compter de la date de priorité, la demande d'examen préliminaire international doit être présentée à l'administration compétente chargée de l'examen préliminaire international avant l'expiration d'un délai de 19 mois à compter de la date de priorité.

Il appartient exclusivement au déposant de veiller au respect du délai de 19 mois.

Il est à noter que seul un déposant qui est ressortissant d'un Etat contractant du PCT lié par le chapitre II ou qui y a son domicile peut présenter une demande d'examen préliminaire international.

### RAPPEL CONCERNANT L'OUVERTURE DE LA PHASE NATIONALE (article 22 ou 39.1))

Si le déposant souhaite que la demande internationale procède en phase nationale, il doit, dans le délai de 20 mois ou de 30 mois, ou plus pour ce qui concerne certains offices, accomplir les actes mentionnés dans ces dispositions auprès de chaque office désigné ou élu.

Pour d'autres informations importantes concernant les délais et les actes à accomplir pour l'ouverture de la phase nationale, voir l'annexe du formulaire PCT/IB/301 (Notification de la réception de l'exemplaire original) et le volume II du Guide du déposant du PCT.

Bureau international de l'OMPI 34, chemin des Colombettes 1211 Genève 20, Suisse	Fonctionnaire autorisé J. Zahra
no de télécopieur (41-22) 740.14.35	no de téléphone (41-22) 338.83.38



Suite du formulaire PCT/IB/308

**AVIS INFORMANT LE DEPOSANT DE LA COMMUNICATION DE  
LA DEMANDE INTERNATIONALE AUX OFFICES DESIGNES**

<b>Date d'expédition (jour/mois/année)</b> 15 mars 2001 (15.03.01)	<b>AVIS IMPORTANT</b>
<b>Référence du dossier du déposant ou du mandataire</b> 341225/18374	<b>Demande internationale no</b> PCT/FR00/02433
<p>Il est notifié au déposant que, au moment de l'établissement du présent avis, le délai fixé à la règle 46.1 pour le dépôt de modifications selon l'article 19 n'était pas encore expiré et que le Bureau international n'avait pas reçu de modifications ni de déclaration l'informant que le déposant ne souhaitait pas présenter de modifications.</p>	



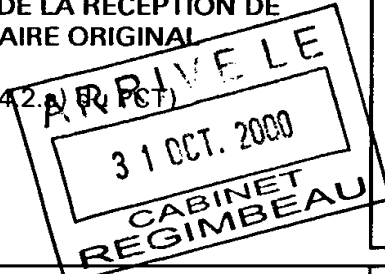
# TRAITE DE COOPERATION EN MATIERE DE BREVETS

Expéditeur: le BUREAU INTERNATIONAL

PCT

NOTIFICATION DE LA RECEPTION DE  
L'EXEMPLAIRE ORIGINAL

(règle 242.2 PCT)



Destinataire:

MARTIN, Jean-Jacques  
Cabinet Regimbeau  
26, avenue Kléber  
F-75116 Paris  
FRANCE

Date d'expédition (jour/mois/année) 20 octobre 2000 (20.10.00)	NOTIFICATION IMPORTANTE
Référence du dossier du déposant ou du mandataire 341225/18374	Demande internationale no PCT/FR00/02433

Il est notifié au déposant que le Bureau international a reçu l'exemplaire original de la demande internationale précisée ci-après.

Nom(s) du ou des déposants et de l'Etat ou des Etats pour lesquels ils sont déposants:

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) (pour tous les Etats désignés sauf US)

WEISSENBACH, Jean etc. (pour US seulement)

Date du dépôt international : 04 septembre 2000 (04.09.00)

Date(s) de priorité revendiquée(s) : 03 septembre 1999 (03.09.99)

Date de réception de l'exemplaire original par le Bureau international : 03 octobre 2000 (03.10.00)

Liste des offices désignés :

EP : AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE  
National : CA,JP,US

## ATTENTION

Le déposant doit soigneusement vérifier les indications figurant dans la présente notification. En cas de divergence entre ces indications et celles que contient la demande internationale, il doit aviser immédiatement le Bureau international.

En outre, l'attention du déposant est appelée sur les renseignements donnés dans l'annexe en ce qui concerne

- ☒ les délais dans lesquels doit être abordée la phase nationale
- ☒ la confirmation des désignations faites par mesure de précaution
- ☐ les exigences relatives aux documents de priorité.

Une copie de la présente notification est envoyée à l'office récepteur et à l'administration chargée de la recherche internationale.

Bureau international de l'OMPI 34, chemin des Colombettes 1211 Genève 20, Suisse n° de télécopieur (41-22) 740.14.35	Fonctionnaire autorisé Jocelyne Rey-Millet n° de téléphone (41-22) 338.83.38
---	--



**RENSEIGNEMENTS CONCERNANT LES DELAIS DANS LESQUELS DOIT ETRE ABORDEE  
LA PHASE NATIONALE**

Il est rappelé au déposant qu'il doit aborder la "phase nationale" auprès de chacun des offices désignés indiqués sur la notification de la réception de l'exemplaire original (formulaire PCT/IB/301) en payant les taxes nationales et en remettant les traductions, telles qu'elles sont prescrites par les législations nationales.

Le délai d'accomplissement de ces actes de procédure est de **20 MOIS** à compter de la date de priorité ou, pour les Etats désignés qui ont été élus par le déposant dans une demande d'examen préliminaire international ou dans une élection ultérieure, de **30 MOIS** à compter de la date de priorité, à condition que cette élection ait été effectuée avant l'expiration du 19<sup>e</sup> mois à compter de la date de priorité. Certains offices désignés (ou élus) ont fixé des délais qui expirent au-delà de 20 ou 30 mois à compter de la date de priorité. D'autres offices accordent une prolongation des délais ou un délai de grâce, dans certains cas moyennant le paiement d'une taxe supplémentaire.

En plus de ces actes de procédure, le déposant devra dans certains cas satisfaire à d'autres exigences particulières applicables dans certains offices. **Il appartient au déposant** de veiller à remplir en temps voulu les conditions requises pour l'ouverture de la phase nationale. La majorité des offices désignés n'envoient pas de rappel à l'approche de la date limite pour aborder la phase nationale.

**Des informations détaillées concernant les actes de procédure à accomplir pour aborder la phase nationale auprès de chaque office désigné, les délais applicables et la possibilité d'obtenir une prolongation des délais ou un délai de grâce et toutes autres conditions applicables figurent dans le volume II du Guide du déposant du PCT. Les exigences concernant le dépôt d'une demande d'examen préliminaire international sont exposées dans le chapitre IX du volume I du Guide du déposant du PCT.**

GR et ES sont devenues liées par le chapitre II du PCT le 7 septembre 1996 et le 6 septembre 1997, respectivement, et peuvent donc être élues dans une demande d'examen préliminaire international ou dans une élection ultérieure présentée le 7 septembre 1996 (ou à une date postérieure) ou le 6 septembre 1997 (ou à une date postérieure), respectivement, quelle que soit la date de dépôt de la demande internationale (voir le second paragraphe, ci-dessus).

Veuillez noter que seul un déposant qui est ressortissant d'un Etat contractant du PCT lié par le chapitre II ou qui y a son domicile peut présenter une demande d'examen préliminaire international.

**CONFIRMATION DES DESIGNATIONS FAITES PAR MESURE DE PRECAUTION**

Seules les désignations expresses faites dans la requête conformément à la règle 4.9.a) figurent dans la présente notification. Il est important de vérifier si ces désignations ont été faites correctement. Des erreurs dans les désignations peuvent être corrigées lorsque des désignations ont été faites par mesure de précaution en vertu de la règle 4.9.b). Toute désignation ainsi faite peut être confirmée conformément aux dispositions de la règle 4.9.c) avant l'expiration d'un délai de 15 mois à compter de la date de priorité. En l'absence de confirmation, une désignation faite par mesure de précaution sera considérée comme retirée par le déposant. Il ne sera adressé aucun rappel ni invitation. Pour confirmer une désignation, il faut déposer une déclaration précisant l'Etat désigné concerné (avec l'indication de la forme de protection ou de traitement souhaitée) et payer les taxes de désignation et de confirmation. La confirmation doit parvenir à l'office récepteur dans le délai de 15 mois.

**EXIGENCES RELATIVES AUX DOCUMENTS DE PRIORITE**

Pour les déposants qui n'ont pas encore satisfait aux exigences relatives aux documents de priorité, il est rappelé ce qui suit.

Lorsque la priorité d'une demande nationale, régionale ou internationale antérieure est revendiquée, le déposant doit présenter une copie de cette demande antérieure, certifiée conforme par l'administration auprès de laquelle elle a été déposée ("document de priorité"), à l'office récepteur (qui la transmettra au Bureau international) ou directement au Bureau international, avant l'expiration d'un délai de 16 mois à compter de la date de priorité, étant entendu que tout document de priorité peut être présenté au Bureau international avant la date de publication de la demande internationale, auquel cas ce document sera réputé avoir été reçu par le Bureau international le dernier jour du délai de 16 mois (règle 17.1.a)).

Lorsque le document de priorité est délivré par l'office récepteur, le déposant peut, au lieu de présenter ce document, demander à l'office récepteur de le préparer et de le transmettre au Bureau international. La requête à cet effet doit être formulée avant l'expiration du délai de 16 mois et peut être soumise au paiement d'une taxe (règle 17.1.b)).

Si le document de priorité en question n'est pas fourni au Bureau international, ou si la demande adressée à l'office récepteur de préparer et de transmettre le document de priorité n'a pas été faite (et la taxe correspondante acquittée, le cas échéant) avant l'expiration du délai applicable mentionné aux paragraphes précédents, tout Etat désigné peut ne pas tenir compte de la revendication de priorité; toutefois, aucun office désigné ne peut décider de ne pas tenir compte de la revendication de priorité avant d'avoir donné au déposant la possibilité de remettre le document de priorité dans un délai raisonnable en l'espèce.

Lorsque plusieurs priorités sont revendiquées, la date de priorité à prendre en considération aux fins du calcul du délai de 16 mois est la date du dépôt de la demande la plus ancienne dont la priorité est revendiquée.



# TRAITE DE COOPERATION EN MATIERE DE BREVETS

Expéditeur : le BUREAU INTERNATIONAL

PCT

## NOTIFICATION RELATIVE A LA PRESENTATION OU A LA TRANSMISSION DU DOCUMENT DE PRIORITE

(instruction administrative 411 du PCT)


Destinataire:

MARTIN, Jean-Jacques  
Cabinet Regimbeau  
26, avenue Kléber  
F-75116 Paris  
FRANCE

Date d'expédition (jour/mois/année) 20 octobre 2000 (20.10.00)	
Référence du dossier du déposant ou du mandataire 341225/18374	<b>NOTIFICATION IMPORTANTE</b>
Demande internationale no PCT/FR00/02433	Date du dépôt international (jour/mois/année) 04 septembre 2000 (04.09.00)
Date de publication internationale (jour/mois/année) Pas encore publiée	Date de priorité (jour/mois/année) 03 septembre 1999 (03.09.99)
Déposant CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) etc	

1. La date de réception (sauf lorsque les lettres "NR" figurent dans la colonne de droite) par le Bureau international du ou des documents de priorité correspondant à la ou aux demandes énumérées ci-après est notifiée au déposant. Sauf indication contraire consistant en un astérisque figurant à côté d'une date de réception, ou les lettres "NR", dans la colonne de droite, le document de priorité en question a été présenté ou transmis au Bureau international d'une manière conforme à la règle 17.1.a) ou b).
2. Ce formulaire met à jour et remplace toute notification relative à la présentation ou à la transmission du document de priorité qui a été envoyée précédemment.
3. Un **astérisque(\*)** figurant à côté d'une date de réception dans la colonne de droite signale un document de priorité présenté ou transmis au Bureau international mais de manière non conforme à la règle 17.1.a) ou b). Dans ce cas, **l'attention du déposant est appelée** sur la règle 17.1.c) qui stipule qu'aucun office désigné ne peut décider de ne pas tenir compte de la revendication de priorité avant d'avoir donné au déposant la possibilité de remettre le document de priorité dans un délai raisonnable en l'espèce.
4. Les **lettres "NR"** figurant dans la colonne de droite signalent un document de priorité que le Bureau international n'a pas reçu ou que le déposant n'a pas demandé à l'office récepteur de préparer et de transmettre au Bureau international, conformément à la règle 17.1.a) ou b), respectivement. Dans ce cas, **l'attention du déposant est appelée** sur la règle 17.1.c) qui stipule qu'aucun office désigné ne peut décider de ne pas tenir compte de la revendication de priorité avant d'avoir donné au déposant la possibilité de remettre le document de priorité dans un délai raisonnable en l'espèce.

<u>Date de priorité</u>	<u>Demande de priorité n°</u>	<u>Pays, office régional ou office récepteur selon le PCT</u>	<u>Date de réception du document de priorité</u>
03 sept 1999 (03.09.99)	99/11097	FR	03 octo 2000 (03.10.00)

<p>Bureau international de l'OMPI 34, chemin des Colombettes 1211 Genève 20, Suisse</p> <p>no de télécopieur (41-22) 740.14.35</p>	<p>Fonctionnaire autorisé:</p> <p>Jocelyne Rey-Millet </p> <p>no de téléphone (41-22) 338.83.38</p>
--	--



## PCT

## RAPPORT DE RECHERCHE INTERNATIONALE

(article 18 et règles 43 et 44 du PCT)

Référence du dossier du déposant ou du mandataire <b>341225/18374</b>	<b>POUR SUITE</b> voir la notification de transmission du rapport de recherche internationale (formulaire PCT/ISA/220) et, le cas échéant, le point 5 ci-après <b>A DONNER</b>	
Demande internationale n° <b>PCT/FR 00/ 02433</b>	Date du dépôt international (jour/mois/année) <b>04/09/2000</b>	(Date de priorité (la plus ancienne) (jour/mois/année) <b>03/09/1999</b>
Déposant  <b>CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS)</b>		

Le présent rapport de recherche internationale, établi par l'administration chargée de la recherche internationale, est transmis au déposant conformément à l'article 18. Une copie en est transmise au Bureau international.

Ce rapport de recherche internationale comprend 3 feuilles.



Il est aussi accompagné d'une copie de chaque document relatif à l'état de la technique qui y est cité.

**1. Base du rapport**

- a. En ce qui concerne la **langue**, la recherche internationale a été effectuée sur la base de la demande internationale dans la langue dans laquelle elle a été déposée, sauf indication contraire donnée sous le même point.



la recherche internationale a été effectuée sur la base d'une traduction de la demande internationale remise à l'administration.

- b. En ce qui concerne les **séquences de nucléotides ou d'acides aminés** divulguées dans la demande internationale (le cas échéant), la recherche internationale a été effectuée sur la base du listage des séquences :



contenu dans la demande internationale, sous forme écrite.



déposée avec la demande internationale, sous forme déchiffrable par ordinateur.



remis ultérieurement à l'administration, sous forme écrite.



remis ultérieurement à l'administration, sous forme déchiffrable par ordinateur.



La déclaration, selon laquelle le listage des séquences présenté par écrit et fourni ultérieurement ne vas pas au-delà de la divulgation faite dans la demande telle que déposée, a été fournie.



La déclaration, selon laquelle les informations enregistrées sous forme déchiffrable par ordinateur sont identiques à celles du listage des séquences présenté par écrit, a été fournie.

2. ☒ **Il a été estimé que certaines revendications ne pouvaient pas faire l'objet d'une recherche** (voir le cadre I).

3. ☐ **Il y a absence d'unité de l'invention** (voir le cadre II).

**4. En ce qui concerne le titre,**

le texte est approuvé tel qu'il a été remis par le déposant.



Le texte a été établi par l'administration et a la teneur suivante:

**5. En ce qui concerne l'abrégé,**

le texte est approuvé tel qu'il a été remis par le déposant



le texte (reproduit dans le cadre III) a été établi par l'administration conformément à la règle 38.2b). Le déposant peut présenter des observations à l'administration dans un délai d'un mois à compter de la date d'expédition du présent rapport de recherche internationale.

**6. La figure des dessins à publier avec l'abrégé est la Figure n°**

suggérée par le déposant.



parce que le déposant n'a pas suggéré de figure.



parce que cette figure caractérise mieux l'invention.



Aucune des figures n'est à publier.



## SUITE DES RENSEIGNEMENTS INDIQUES SUR PCT/ISA/ 210

## Suite du cadre I.2

Les revendications 13-15, 19-36 concernent en partie des séquences d'acides nucléiques promotrices et/ou régulatrices de l'expression du gène SPG4, ainsi que des polypéptides correspondants, sans y donner une véritable caractérisation technique. Ces revendications comprennent donc aussi des nucléiques et polypeptides dérivés de tout autre gène contrôlant SPG4, ce qui n'est ni défini ni caractérisé dans la description. En conséquence, ces revendications sont ambiguës et vagues, et leur sujet-matière n'est pas suffisamment décrit et démontré en accord avec les Articles 5 et 6 TCP. Aucune recherche ne peut être établie pour des revendications spéculatives dont l'objet ne serait qu'une simple liste de résultats à obtenir dans le futur.

L'attention du déposant est attirée sur le fait que les revendications, ou des parties de revendications, ayant trait aux inventions pour lesquelles aucun rapport de recherche n'a été établi ne peuvent faire obligatoirement l'objet d'un rapport préliminaire d'examen (Règle 66.1(e) PCT). Le déposant est averti que la ligne de conduite adoptée par l'OEB agissant en qualité d'administration chargée de l'examen préliminaire international est, normalement, de ne pas procéder à un examen préliminaire sur un sujet n'ayant pas fait l'objet d'une recherche. Cette attitude restera inchangée, indépendamment du fait que les revendications aient ou n'aient pas été modifiées, soit après la réception du rapport de recherche, soit pendant une quelconque procédure sous le Chapitre II.



## A. CLASSEMENT DE L'OBJET DE LA DEMANDE

CIB 7 C12N15/12 C12N5/10 C07K14/47 C07K16/18 C12Q1/68  
G01N33/50 A01K67/027

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

## B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)

CIB 7 C12N C07K A01K G01N C12Q

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si réalisable, termes de recherche utilisés)

EMBL, STRAND, BIOSIS

## C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie °	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	<p>KIKUNO R. ET AL.: "Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 cDNA clones from brain which code for large proteins in vitro." DNA RES., vol. 6, 30 juin 1999 (1999-06-30), pages 197-205, XP000852618 le document en entier -&amp; DATABASE EMBL 'en ligne! Accession No. AB029006, 4 août 1999 (1999-08-04) OHARA O.: "Homo sapiens mRNA for KIAA1083 protein" XP002140994 le document en entier</p> <p style="text-align: center;">--- -/--</p>	<p>1,4,5, 14,16, 18-21, 24-30</p>

☒ Voir la suite du cadre C pour la fin de la liste des documents

☐ Les documents de familles de brevets sont indiqués en annexe

## ° Catégories spéciales de documents cités:

\*A\* document définissant l'état général de la technique, non considéré comme particulièrement pertinent

\*E\* document antérieur, mais publié à la date de dépôt international ou après cette date

\*L\* document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)

\*O\* document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens

\*P\* document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

\*T\* document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention

\*X\* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

\*Y\* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

\*&\* document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

4 janvier 2001

Date d'expédition du présent rapport de recherche internationale

25/01/2001

Nom et adresse postale de l'administration chargée de la recherche internationale

Office Européen des Brevets, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Fonctionnaire autorisé

Galli, I



## C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	<p>DATABASE EMBL SEQUENCES 'en ligne!  Accession No. HS1179930,  29 mars 1997 (1997-03-29)  MYERS R.M.: "human STS SHGC-44567"  XP002156510  brin complémentaire aux nucléotides  2204-2575 de la séquence no. 2</p> <p>----</p>	<p>1,4,5,  14,16,  18-21,  24-30</p>
P,X	<p>HAZAN J. ET AL.: "Spastin, a new AAA  protein, is altered in the most frequent  form of autosomal dominant spastic  paraplegia"  NAT. GENET.,  vol. 23, novembre 1999 (1999-11), pages  296-303, XP000914979  le document en entier</p> <p>----</p>	<p>1-36</p>
A	<p>HEINZLEF O. ET AL.: "Mapping of a  complicated familial spastic paraplegia to  locus SPG4 on chromosome 2p"  J. MED. GENET.,  vol. 35, no. 2, février 1998 (1998-02),  pages 89-93, XP000914971  cité dans la demande  le document en entier</p> <p>-----</p>	<p>1-36</p>



(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION  
EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété  
Intellectuelle  
Bureau international



(43) Date de la publication internationale  
15 mars 2001 (15.03.2001)

PCT

(10) Numéro de publication internationale  
**WO 01/18198 A1**

(51) Classification internationale des brevets<sup>7</sup>: C12N 15/12,  
5/10, C07K 14/47, 16/18, C12Q 1/68, G01N 33/50, A01K  
67/027

F-75015 Paris (FR). HAZAN, Jamilé [FR/FR]; 22, rue  
des Cordelières, F-75013 Paris (FR).

(21) Numéro de la demande internationale:  
PCT/FR00/02433

(74) Mandataires: MARTIN, Jean-Jacques etc.; Cabinet  
Regimbeau, 26, avenue Kléber, F-75116 Paris (FR).

(22) Date de dépôt international:  
4 septembre 2000 (04.09.2000)

(81) États désignés (national): CA, JP, US.

(25) Langue de dépôt: français

(84) États désignés (régional): brevet européen (AT, BE, CH,  
CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT,  
SE).

(26) Langue de publication: français

(30) Données relatives à la priorité:  
99/11097 3 septembre 1999 (03.09.1999) FR

**Publiée:**

- Avec rapport de recherche internationale.
- Avant l'expiration du délai prévu pour la modification des  
revendications, sera republiée si des modifications sont  
reçues.

(71) Déposant (pour tous les États désignés sauf US):  
CENTRE NATIONAL DE LA RECHERCHE SCI-  
ENTIFIQUE (CNRS) [FR/FR]; 3, rue Michel-Ange,  
F-75794 Paris Cedex 16 (FR).

En ce qui concerne les codes à deux lettres et autres abrégia-  
tions, se référer aux "Notes explicatives relatives aux codes et  
abréviations" figurant au début de chaque numéro ordinaire de  
la Gazette du PCT.

(72) Inventeurs; et

(75) Inventeurs/Déposants (pour US seulement): WEIS-  
SENBACH, Jean [FR/FR]; 163, rue de Vaugirard,

(54) Title: CLONING, EXPRESSION AND CHARACTERISATION OF THE SPG4 GENE RESPONSIBLE FOR THE MOST  
FREQUENT FORM OF AUTOSOMAL SPASTIC PARAPLEGIA

(54) Titre: CLONAGE, EXPRESSION ET CARACTERISATION DU GENE SPG4 RESPONSABLE DE LA FORME LA PLUS  
FREQUENTE DE PARAPLEGIE SPASTIQUE AUTOSOMIQUE DOMINANTE

(57) Abstract: The invention concerns the identification and characterisation of the SPG4 gene coding for spastin, and some mu-  
tations thereof responsible for the most frequent form of autosomal dominant familial spastic paraplegia, the cloning and the char-  
acterisation of its cDNA and the corresponding polypeptides. The invention also concerns vectors, transformed cells and transgenic  
animals as well as diagnostic methods and kits, and methods for selecting a chemical or biological compound capable of directly or  
indirectly interacting with said polypeptide.

(57) Abrégé: L'invention concerne l'identification et la caractérisation du gène SPG4 codant pour la spastin, et certaines de ses  
mutations responsables de la forme la plus fréquente de paraplégie spastique familiale (PSF) autosomique dominante, le clonage et  
la caractérisation de son ADNc ainsi que les polypeptides correspondants. L'invention concerne également des vecteurs, des cellules  
transformées et des animaux transgéniques ainsi que des méthodes et trousse de diagnostic, et des méthodes de sélection d'un  
composé chimique ou biochimique capable d'interagir directement ou indirectement avec un polypeptide selon l'invention.

WO 01/18198 A1



!

.

.

.

.

CLONAGE, EXPRESSION ET CARACTERISATION DU GENE SPG4 RESPONSABLE DE LA FORME LA PLUS FREQUENTE DE PARAPLEGIE SPASTIQUE AUTOSOMIQUE DOMINANTE.

5 L'invention concerne l'identification et la caractérisation du gène SPG4 codant pour la spastin, responsable de la forme la plus fréquente de paraplégie spastique familiale (PSF) autosomique dominante, le clonage et la caractérisation de son ADNc ainsi que les polypeptides correspondants. L'invention concerne également des vecteurs, des cellules transformées et des animaux transgéniques ainsi que des  
10 méthodes et trousse de diagnostic, et des méthodes de sélection d'un composé chimique ou biochimique capable d'interagir directement ou indirectement avec un polypeptide selon l'invention.

Les paraplégies spastiques familiales (PSF) sont des affections dégénératives du système nerveux central caractérisées par une spasticité bilatérale et progressive  
15 des membres inférieurs. Elles se manifestent cliniquement par des difficultés de la marche pouvant évoluer en une paralysie totale des deux jambes. La physiopathologie de cet ensemble de maladies est à ce jour peu documentée ; toutefois, les données anatomopathologiques permettent de conclure que l'atteinte est limitée aux faisceaux pyramidaux responsables de la motricité volontaire dans la moelle épinière (Reid,  
20 1997). Il existe différentes formes cliniques et génétiques de PSF. On distingue cliniquement les PSF dites « pures » correspondant à une spasticité isolée des membres inférieurs, des PSF « complexes » pour lesquelles la spasticité des jambes est associée à d'autres manifestations cliniques de type neurologique ou non (Bruyn et al., 1991). D'un point de vue génétique, les PSF peuvent être transmises selon le  
25 mode autosomique dominant (PSF-AD), autosomique récessif (PSF-AR) ou lié au chromosome X (PSF-X). La forme « pure » de PSF qui est le plus souvent transmise selon le mode autosomique dominant demeure la plus fréquente (environ 80 % des PSF) (Reid, 1997). L'incidence des PSF qui reste difficile à estimer en raison des rares études épidémiologiques et de l'importante variabilité clinique varie de 0,9 : 100 000 au  
30 Danemark, 3 à 9,6 : 100 000 dans certaines régions d'Espagne (Polo et al., 1991) ou 14 : 100 000 en Norvège (Skre, 1974) (environ 3 : 100 000 en France).

En plus de cette grande variabilité clinique qui est observée non seulement entre les différentes familles mais aussi entre différents membres atteints d'une même famille, les PSF se caractérisent également par une importante hétérogénéité  
35 génétique. Dans le cas des PSF-AD, quatre loci ont été identifiés à ce jour sur les

chromosomes 14 (locus SPG3) (Hazan et al., 1993), 2 (locus SPG4) (Hazan et al., 1994 ; Hentali et al., 1994), 15 (locus SPG6) (Fink et al., 1995) et 8 (locus SPG8) (Hedera et al., 1999). L'étude d'un grand nombre de familles présentant une PSF-AD a montré que le gène porté par le chromosome 2 est un locus majoritaire de cette forme de la maladie, retrouvé dans 40 à 50 % des familles analysées (The Hereditary Spastic Paraplegia Working Group, 1996 ; Durr et al., 1996). Un phénomène d'anticipation a été observé dans certaines familles de PSF liée au locus SPG4 ; ce phénomène a par la suite été associé à l'expansion d'une répétition (CAG)<sub>n</sub> mise en évidence dans 6 familles danoises (Nielsen et al., 1997) par la technique de RED (pour Rapid Expansion Detection). Cette expansion n'a cependant jamais pu être confirmée dans aucune des familles testées par cette même méthode ni par la recherche systématique de séquences de type (CAG)<sub>n</sub> dans les cartes physiques composées de clones YAC (pour Yeast Artificial Chromosome) ou BAC (pour Bacterial Artificial Chromosome) (Hazan et al., Genomics, 60 (3), 309-19, 1999).

A ce jour, trois gènes responsables de deux formes de PSF-X et d'une forme de PSF-AR ont été identifiés. Des mutations dans le gène, codant pour une molécule d'adhérence cellulaire spécifique des neurones, L1-CAM (pour L1 Cell Adhesion Molecule), et localisé en Xq28 (locus SPG1) causent une forme complexe de PSF (Jouet et al., 1994) dans laquelle la spasticité est associée à un retard mental, alors que des mutations dans le gène PLP (pour ProteoLipid Protein) situé en Xq21 (locus SPG2), codant pour une molécule constitutive de la couche de myéline, sont à l'origine de formes pures et complexes de PSF-X (Saugier-Veber, P. et al., 1994). Plus récemment des mutations dans un gène localisé en 16q24.3 (locus SPG7), qui code pour la paraplegin, une ATPase mitochondriale de la famille protéique des AAA (pour «ATPases Associated with diverse cellular Activities») (Confalonieri et al., 1995) ont été associées à des formes complexes et pures de PSF-AR (Casari et al., 1998).

Ainsi, il reste aujourd'hui un grand besoin d'identifier et de caractériser le gène responsable de la forme la plus fréquente de PSF-AD. L'identification de ce gène devrait en particulier permettre, outre la possibilité d'un test de dépistage anténatal chez les familles concernées, de mieux comprendre certains des mécanismes moléculaires engendrant ces dégénérescences spécifiques de faisceaux nerveux de la moelle épinière, voire d'apporter des éléments de réponse quant à un traitement thérapeutique des malades.

Ceci est just ment l'objet de la présente invention.

Après avoir délimité l'intervalle de localisation entre les marqueurs génétiques D2S352 et D2S2347 par l'étude des événements de recombinaison dans les familles de PSF liées au locus SPG4, les inventeurs ont établi un contig de BACs recouvrant une distance physique évaluée à environ 1,5 Mb et entrepris une stratégie de clonage positionnel basée sur le séquençage de l'intervalle SPG4 afin d'identifier de façon exhaustive tous les gènes localisés dans la région candidate. L'analyse de la séquence des deux BACs, D (b336P14) et G (B763N4), a révélé la présence d'un gène composé de 17 exons, s'étendant sur une distance d'environ 100 kb, et présentant une homologie avec les gènes codant pour des protéines de la famille des AAA. La comparaison de la séquence de ce gène entre les individus sains et atteints des familles de PSF-AD a permis de mettre en évidence différentes mutations chez les patients.

L'invention a ainsi pour objet l'identification et la caractérisation du gène SPG4 (ou SPAST) codant pour un nouveau membre nucléaire de la famille des AAA, responsable de la forme la plus fréquente de PSF-AD.

Sous un premier aspect, la présente invention a pour objet un acide nucléique purifié ou isolé du gène SPG4, caractérisé en ce qu'il comprend au moins 15 nucléotides consécutifs, de préférence 20, 25, 30, 35, 40, 45, 50, 75, 100 ou 200 nucléotides consécutifs, d'une séquence choisie parmi le groupe comprenant :

- la séquence SEQ ID No. 1, séquence génomique du gène SPG4 humain ;
- les séquences nucléiques homologues ou variantes de l'acide nucléique de séquence SEQ ID No. 1 ;
- leur séquence complémentaire ; et
- la séquence de leur ARN correspondant.

La présente invention concerne, bien entendu, aussi bien les séquences ADN qu'ARN ainsi que les séquences qui s'hybrident avec elles, de même que les ADNs double brin correspondants.

Par acide nucléique, séquence nucléique ou d'acide nucléique, polynucléotide, oligonucléotide, séquence de polynucléotide, séquence nucléotidique, termes qui seront employés indifféremment dans la présente description, on entendra désigner aussi bien un ADN double brin, un ADN simple brin que des produits de transcription desdits ADNs, et/ou un fragment d'ARN, lesdits fragments naturels isolés, ou de synthèse, comportant ou non des nucléotides non naturels, désignant un enchaînement précis de nucléotides, modifiés ou non, permettant de définir un fragment ou une région d'un acide nucléique. On entend par fragment d'ADN et/ou

d'ARN naturel isolé, ou de synthèse, comportant ou non des nucléotides non naturels, un enchaînement précis de nucléotides, modifiés ou non, permettant de définir un fragment, un segment ou une région d'un acide nucléique.

Il doit être compris que la présente invention ne concerne pas les séquences  
5 nucléotidiques génomiques dans leur environnement chromosomique naturel, c'est-à-dire à l'état naturel. Il s'agit de séquences qui ont été isolées et/ou purifiées, c'est-à-dire qu'elles ont été prélevées directement ou indirectement, par exemple par copie, leur environnement ayant été au moins partiellement modifié.

Par «séquence nucléique homologue», on entendra désigner les séquences  
10 nucléiques présentant, par rapport à la séquence nucléique de référence certaines modifications comme en particulier une délétion, une troncation, un allongement, une fusion chimérique, et/ou une mutation, notamment ponctuelle, et dont la séquence nucléique présente au moins 80 %, de préférence 90 % ou 95 %, d'identité après alignement avec la séquence nucléique de référence.

15 Par « pourcentage d'identité » entre deux séquences d'acide nucléique ou d'acides aminés au sens de la présente invention, on entend désigner un pourcentage de nucléotides ou de résidus d'acides aminés identiques entre les deux séquences à comparer, obtenu après le meilleur alignement, ce pourcentage étant purement statistique et les différences entre les deux séquences étant réparties au hasard et sur  
20 toute leur longueur. Les comparaisons de séquences entre deux séquences d'acide nucléique ou d'acides aminés sont traditionnellement réalisées en comparant ces séquences après les avoir alignées de manière optimale, ladite comparaison étant réalisée par segment ou par « fenêtre de comparaison » pour identifier et comparer les régions locales de similarité de séquence. L'alignement optimal des séquences pour la  
25 comparaison peut être réalisé, outre manuellement, au moyen de l'algorithme d'homologie locale de Smith et Waterman (1981) [Ad. App. Math. 2:482], au moyen de l'algorithme d'homologie locale de Needleman et Wunsch (1970) [J. Mol. Biol. 48:443], au moyen de la méthode de recherche de similarité de Pearson et Lipman (1988) [Proc. Natl. Acad. Sci. USA 85:2444], au moyen de logiciels informatiques utilisant ces  
30 algorithmes (GAP, BESTFIT, FASTA et TFASTA dans le Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI, ou encore par les logiciels de comparaison BLAST N ou BLAST P).

Le pourcentage d'identité entre deux séquences d'acide nucléique ou d'acides aminés est déterminé en comparant ces deux séquences alignées de manière  
35 optimale par fenêtre de comparaison dans laquelle la région de la séquence d'acide

nucléique ou d'acides aminés à comparer peut comprendre des additions ou des délétions par rapport à la séquence de référence pour un alignement optimal entre ces deux séquences. Le pourcentage d'identité est calculé en déterminant le nombre de positions identiques pour lesquelles le nucléotide ou le résidu d'acide aminé est  
5 identique entre les deux séquences, en divisant ce nombre de positions identiques par le nombre total de positions dans la fenêtre de comparaison et en multipliant le résultat obtenu par 100 pour obtenir le pourcentage d'identité entre ces deux séquences.

Par exemple, on pourra utiliser le programme BLAST, « BLAST 2 sequences » (Tatusova et al., "Blast 2 sequences - a new tool for comparing protein and nucleotide  
10 sequences", FEMS Microbiol. Lett. 174:247-250) disponible sur le site <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>, les paramètres utilisés étant ceux donnés par défaut (en particulier pour les paramètres « open gap pénalité » : 5, et « extension gap pénalité » : 2 ; la matrice choisie étant par exemple la matrice « BLOSUM 62 » proposée par le programme), le pourcentage d'identité entre les deux séquences à  
15 comparer étant calculé directement par le programme.

Il s'agit de préférence de séquences dont les séquences complémentaires sont susceptibles de s'hybrider spécifiquement avec une des séquences de l'invention. De préférence, les conditions d'hybridation spécifiques ou de forte stringence seront telles qu'elles assurent au moins 80 %, de préférence 90 % ou 95 % d'identité après  
20 alignement entre l'une des deux séquences et la séquence complémentaire de l'autre.

Une hybridation dans des conditions de forte stringence signifie que les conditions de température et de force ionique sont choisies de telle manière qu'elles permettent le maintien de l'hybridation entre deux fragments d'ADN complémentaires. A titre illustratif, des conditions de forte stringence de l'étape d'hybridation aux fins de  
25 définir les fragments polynucléotidiques décrits ci-dessus, sont avantageusement les suivantes.

L'hybridation ADN-ADN ou ADN-ARN est réalisée en deux étapes : (1) préhybridation à 42°C pendant 3 heures en tampon phosphate (20 mM, pH 7,5) contenant 5 x SSC (1 x SSC correspond à une solution 0,15 M NaCl + 0,015 M citrate  
30 de sodium), 50 % de formamide, 7 % de sodium dodécyl sulfate (SDS), 10 x Denhardt's, 5 % de dextran sulfate et 1 % d'ADN de sperme de saumon ; (2) hybridation proprement dite pendant 20 heures à une température dépendant de la taille de la sonde (i.e. : 42°C, pour une sonde de taille > 100 nucléotides) suivie de 2 lavages de 20 minutes à 20°C en 2 x SSC + 2 % SDS, 1 lavage de 20 minutes à 20°C  
35 en 0,1 x SSC + 0,1 % SDS. Le dernier lavage est pratiqué en 0,1 x SSC + 0,1 % SDS

pendant 30 minutes à 60°C pour une sonde de taille > 100 nucléotides. Les conditions d'hybridation de forte stringence décrites ci-dessus pour un polynucléotide de taille définie, seront adaptées par l'homme du métier pour des oligonucléotides de taille plus grande ou plus petite, selon l'enseignement de Sambrook et al., 1989.

5 Par «séquence nucléique variante» ou «acide nucléique variant» d'une séquence nucléique de référence, on entendra désigner l'ensemble des séquences nucléiques correspondant à des variants alléliques, c'est-à-dire des variations individuelles de la séquence nucléique de référence. Ces séquences mutées naturelles  
10 correspondent à des polymorphismes présents chez les mammifères, en particulier chez l'être humain et, notamment, à des polymorphismes pouvant conduire à la survenue et/ou au développement de pathologie.

Si les séquences selon l'invention concernent les séquences normales, elles concernent également les séquences mutées dans la mesure où elles comportent au moins une mutation ponctuelle et de préférence au plus 10 % de mutations par rapport  
15 à la séquence normale.

En particulier, les séquences nucléiques variantes comprendront toute séquence d'au moins 15 nucléotides consécutifs, de préférence 20, 25, 30, 50, 100 ou 200 nucléotides consécutifs, d'une séquence polymorphique de la séquence génomique du gène SPG4 humain de séquence SEQ ID No. 1, et dont la séquence  
20 d'acide nucléique présente par rapport à la séquence SEQ ID No. 1 au moins une mutation correspondant notamment à une troncation, délétion, substitution et/ou addition de résidu d'acide aminé. Dans le cas présent, les séquences nucléiques variantes présentant au moins une mutation seront ici liées aux pathologies de type PSF-AD liées au locus SPG4.

25 De préférence, la présente invention concerne des séquences nucléiques mutées dans lesquelles les mutations conduisent à une modification de la séquence d'acides aminés du polypeptide codé par la séquence normale.

On entendra également désigner par séquences nucléiques variantes tout ARN ou ADNc résultant d'une mutation d'un site d'épissage de la séquence nucléique  
30 génomique SEQ ID No. 1.

De préférence, l'invention concerne un acide nucléique purifié ou isolé du gène SPG4 selon l'invention, caractérisé en ce qu'il comprend une séquence choisie parmi le groupe comprenant :

a) la séquence SEQ ID No. 1, la séquence SEQ ID No. 2, la séquence SEQ ID No. 72,  
35 la séquence SEQ ID No. 106 ou la séquence d'au moins 15, de préférence 20, 25,

- 30, 35, 40, 45, 50, 75, 100 ou 200 nucléotides consécutifs de la séquence SEQ ID No. 1, SEQ ID No. 2, SEQ ID No. 72 ou SEQ ID No. 106 ;
- b) les séquences nucléiques homologues ou variantes des séquences SEQ ID No. 1, SEQ ID No. 2, SEQ ID No. 72 ou SEQ ID No. 106 ; et
- 5 c) la séquence complémentaire ou la séquence de l'ARN correspondant aux séquences telles que définies en a) et b),
- référence à l'exception de l'acide nucléique identifié dans la banque de données Bank sous le numéro d'accèsion AB029006.

L'acide nucléique dont la séquence est divulguée dans la banque de données GenBank sous le numéro d'accèsion AB029006, correspond à la séquence de l'un des 100 ADNc issus d'une banque d'ARNm de cerveau humain identifiée par le Kazusa DNA Research Institute au Japon (Kikuno et al., DNA Research, 6, 197-205, 1999).

- De préférence, l'invention concerne un acide nucléique purifié ou isolé selon
- 15 l'invention, caractérisé en ce qu'il comprend au moins une séquence d'au moins 15 nucléotides consécutifs, de préférence 20, 25, 30, 50 ou 75 nucléotides consécutifs du fragment nt 714-809, extrémités incluses, de la séquence SEQ ID No. 2, de sa séquence complémentaire ou de la séquence de son ARN correspondant.

- L'invention concerne de préférence un acide nucléique purifié ou isolé selon la
- 20 présente invention, caractérisé en ce qu'il comprend une séquence choisie parmi le groupe suivant :
- la séquence SEQ ID No. 1 ;
  - la séquence SEQ ID No. 2, séquence de l'ADNc codant pour la spastin humaine ;
  - les séquences SEQ ID No. 72 et SEQ ID No. 106, la séquence SEQ ID No. 72
  - 25 représentant la séquence de l'ADNc incomplet codant pour la spastin murine représentée à la figure 5, ligne "Mouse" et la SEQ ID No. 106 sa séquence complète ;
  - les séquences nucléiques homologues ou variantes des séquences SEQ ID No. 1, SEQ ID No. 2, SEQ ID No. 72 ou SEQ ID No. 106 ;
  - leur séquence complémentaire ; et
  - 30 - la séquence de leur ARN correspondant.

De préférence, l'invention concerne un acide nucléique purifié ou isolé selon l'invention, caractérisé en ce qu'il comprend au moins une mutation correspondant à un polymorphisme naturel chez l'Homme, notamment dont la position et la nature sont identifiées dans le tableau 5.

Les amorces ou sondes, caractérisées en ce qu'elles comprennent une séquence d'un acide nucléique selon l'invention, font également partie de l'invention.

La présente invention concerne ainsi l'ensemble des amorces qui peuvent être déduites des séquences nucléotidiques de l'invention et qui peuvent permettre de  
5 mettre en évidence lesdites séquences nucléotidiques de l'invention, en particulier les séquences mutées, en utilisant notamment une méthode d'amplification telle que la méthode PCR, ou une méthode apparentée.

La présente invention concerne également l'ensemble des sondes qui peuvent être déduites des séquences nucléotidiques de l'invention, notamment des séquences  
10 capables de s'hybrider avec elles, et qui peuvent permettre de mettre en évidence lesdites séquences nucléotidiques, en particulier de discriminer les séquences normales des séquences mutées.

La présente invention concerne en particulier les sondes ou amorces de séquences choisies parmi les séquences SEQ ID No. 4 à SEQ ID No. 71.

15 L'invention concerne également l'utilisation d'une séquence d'acide nucléique selon l'invention comme sonde ou amorce, pour la détection, l'identification, le dosage ou l'amplification de séquence d'acide nucléique.

Selon l'invention, les polynucléotides pouvant être utilisés comme sonde ou comme amorce dans des procédés de détection, d'identification, de dosage ou  
20 d'amplification de séquence nucléique, présenteront une taille minimale de 15 bases, de préférence de 20 bases, ou mieux de 25 à 30 bases.

L'ensemble des sondes et amorces selon l'invention pourront être marquées directement ou indirectement par un composé radioactif ou non radioactif par des méthodes bien connues de l'homme du métier, afin d'obtenir un signal détectable et/ou  
25 quantifiable.

Les séquences de polynucléotides selon l'invention non marquées peuvent être utilisées directement comme sonde ou amorce.

Les séquences sont généralement marquées pour obtenir des séquences utilisables pour de nombreuses applications. Le marquage des amorces ou des sondes  
30 selon l'invention est réalisé par des éléments radioactifs ou par des molécules non radioactives.

Parmi les isotopes radioactifs utilisés, on peut citer le  $^{32}\text{P}$ , le  $^{33}\text{P}$ , le  $^{35}\text{S}$ , le  $^3\text{H}$  ou le  $^{125}\text{I}$ . Les entités non radioactives sont sélectionnées parmi les ligands tels la biotine, l'avidine, la streptavidine, la dioxygénine, les haptènes, les colorants, les agents

luminescents tels que les agents radioluminescents, chémiluminescents, bioluminescents, fluorescents, phosphorescents.

Les polynucléotides selon l'invention peuvent ainsi être utilisés comme amorce et/ou sonde dans des procédés mettant en oeuvre notamment la technique de PCR (réaction en chaîne à la polymérase) (Erich, 1989 ; Innis et al., 1990, et Rolfs et al., 1991). Cette technique nécessite le choix de paires d'amorces oligonucléotidiques encadrant le fragment qui doit être amplifié. On peut, par exemple, se référer à la technique décrite dans le brevet américain U.S. N° 4,683,202. Les fragments amplifiés peuvent être identifiés, par exemple après une électrophorèse en gel d'agarose ou de polyacrylamide, ou après une technique chromatographique comme la filtration sur gel ou la chromatographie échangeuse d'ions, puis séquencés. La spécificité de l'amplification peut être contrôlée en utilisant comme amorce les séquences nucléotidiques de polynucléotides de l'invention comme matrice, des plasmides contenant ces séquences ou encore les produits d'amplification dérivés. Les fragments nucléotidiques amplifiés peuvent être utilisés comme réactifs dans des réactions d'hybridation afin de mettre en évidence la présence, dans un échantillon biologique, d'un acide nucléique cible de séquence complémentaire à celle desdits fragments nucléotidiques amplifiés.

L'invention vise également les acides nucléiques susceptibles d'être obtenus par amplification à l'aide d'amorces selon l'invention.

D'autres techniques d'amplification de l'acide nucléique cible peuvent être avantageusement employées comme alternative à la PCR (PCR-like) à l'aide de couple d'amorces de séquences nucléotidiques selon l'invention. Par PCR-like on entendra désigner toutes les méthodes mettant en oeuvre des reproductions directes ou indirectes des séquences d'acides nucléiques, ou bien dans lesquelles les systèmes de marquage ont été amplifiés, ces techniques sont bien entendu connues, en général il s'agit de l'amplification de l'ADN par une polymérase ; lorsque l'échantillon d'origine est un ARN il convient préalablement d'effectuer une transcription reverse. Il existe actuellement de très nombreux procédés permettant cette amplification, comme par exemple la technique SDA (Strand Displacement Amplification) ou technique d'amplification à déplacement de brin (Walker et al., 1992), la technique TAS (Transcription-based Amplification System) décrite par Kwoh et al. en 1989, la technique 3SR (Self-Sustained Sequence Replication) décrite par Guatelli et al. en 1990, la technique NASBA (Nucleic Acid Sequence Based Amplification) décrite par Kievitis et al. en 1991, la technique TMA (Transcription Mediated Amplification), la

technique LCR (Ligase Chain Reaction) décrite par Landegren et al. en 1988 et perfectionnée par Barany et al. en 1991, qui emploie une ligase thermostable, la technique de RCR (Repair Chain Reaction) décrite par Segev en 1992, la technique CPR (Cycling Probe Reaction) décrite par Duck et al. en 1990, la technique  
5 d'amplification à la Q-béta-réplacase décrite par Miele et al. en 1983 et perfectionnée notamment par Chu et al. en 1986 et Lizardi et al. en 1988, puis par Burg et al. ainsi que par Stone et al. en 1996.

Dans le cas où le polynucléotide cible à détecter est un ARNm, on utilisera avantageusement, préalablement à la mise en oeuvre d'une réaction d'amplification à l'aide des amorces selon l'invention ou à la mise en oeuvre d'un procédé de détection à l'aide des sondes de l'invention, une enzyme de type transcriptase reverse afin d'obtenir un ADNc à partir de l'ARNm contenu dans l'échantillon biologique. L'ADNc obtenu servira alors de cible pour les amorces ou les sondes mises en oeuvre dans le  
procédé d'amplification ou de détection selon l'invention.

15 La technique d'hybridation de sondes peut être réalisée de manières diverses (Matthews et al., 1988). La méthode la plus générale consiste à immobiliser l'acide nucléique extrait des cellules de différents tissus ou de cellules en culture sur un support (tels que la nitrocellulose, le nylon, le polystyrène) et à incubé, dans des conditions bien définies, l'acide nucléique cible immobilisé avec la sonde. Après  
20 l'hybridation, l'excès de sonde est éliminé et les molécules hybrides formées sont détectées par la méthode appropriée (mesure de la radioactivité, de la fluorescence ou de l'activité enzymatique liée à la sonde).

Selon un autre mode de mise en oeuvre des sondes nucléiques selon l'invention, ces dernières peuvent être utilisées comme sonde de capture. Dans ce cas,  
25 une sonde, dite « sonde de capture », est immobilisée sur un support et sert à capturer par hybridation spécifique l'acide nucléique cible obtenu à partir de l'échantillon biologique à tester et l'acide nucléique cible est ensuite détecté grâce à une seconde sonde, dite « sonde de détection », marquée par un élément facilement détectable.

Les séquences selon la présente invention de site accepteur ou donneur  
30 d'épissage identifiées au tableau 3 (séquences SEQ ID No. 74 à SEQ ID No. 105) font également partie de la présente invention.

Sous un autre aspect, l'invention comprend une méthode pour le criblage de banques d'ADNc ou d'ADN génomique, ou pour le clonage d'ADNc ou génomique isolé codant pour la spastin, caractérisée en ce qu'elle met en oeuvre une séquence  
35 nucléique selon l'invention.

Parmi ces méthodes, on peut citer notamment :

- le criblage de banques d'ADNc et le clonage des ADNc isolés (Sambrook et al., 1989 ; Suggs et al., 1981 ; Woo et al., 1979), à l'aide des séquences nucléiques selon l'invention ;
- 5 - le criblage de banques génomiques, par exemple de BACs (Chumakov et al., 1992 ; Chumakov et al., 1995) et éventuellement une analyse génétique en FISH (Cherif et al., 1990) à l'aide de séquences selon l'invention, permettant l'isolement et la localisation chromosomique, puis le séquençage complet du gène SPG4 codant pour la spastin.

10 En particulier, ces méthodes selon l'invention pourront être mises en œuvre pour l'identification et ainsi l'obtention de la séquence génomique ou de l'ADNc du gène SPG4 chez d'autres mammifères, notamment la souris.

Ces méthodes de criblage et/ou de clonage comprendront en particulier une étape d'hybridation d'un acide nucléique selon l'invention avec un acide nucléique  
15 contenu dans une banque génomique ou d'ADNc.

L'invention comprend aussi une méthode d'identification des séquences d'acide nucléique promotrices et/ou régulatrices de l'expression du gène SPG4 de séquence SEQ ID No. 1, caractérisée en ce qu'elle met en œuvre un acide nucléique selon l'invention.

20 Les outils informatiques à la disposition de l'homme du métier lui permettent aisément d'identifier à partir des séquences nucléiques génomiques selon l'invention les boîtes régulatrices promotrices nécessaires et suffisantes au contrôle de l'expression génique, notamment les boîtes TATA, CCAAT, GC, ainsi que les séquences régulatrices stimulatrices (« enhancer ») ou inhibitrices (« silencers ») qui  
25 contrôlent en CIS l'expression des gènes selon l'invention ; parmi ces séquences régulatrices, il convient de citer l'IRE, MRE, CRE.

L'invention concerne également des méthodes pour l'identification de mutations portées par le gène SPG4 humain caractérisées en ce qu'elles mettent en œuvre une séquence nucléique selon l'invention, notamment de mutations responsables de la  
30 paraplégie spastique familiale autosomique dominante.

Ces méthodes d'identification de ces mutations comprendront en particulier les étapes suivantes : (i) isolement de l'ADN à partir de l'échantillon biologique à analyser, ou obtention d'un ADNc à partir de l'ARNm de l'échantillon biologique ; (ii) amplification spécifique de l'ADN cible susceptible de présenter une mutation à l'aide d'amorces

sel n l'invention ; (iii) analyse des produits d'amplification, notamment la taille et/ou la séquence des produits d'amplification, par rapport à une séquence de référence.

Par méthodes d'identification de mutation selon l'invention, on entend également désigner une méthode permettant d'obtenir l'acide nucléique sur lequel a  
5 été identifiée ladite mutation.

Font également partie de l'invention, les séquences promotrices et/ou régulatrices du gène SPG4 selon l'invention présentant des mutations susceptibles de modifier l'expression de la protéine correspondante.

Les acides nucléiques caractérisés en ce qu'ils sont susceptibles d'être obtenus  
10 par une des méthodes précédentes selon l'invention, ou les acides nucléiques capables de s'hybrider dans des conditions de forte stringence (homologie d'au moins 80 % entre une des deux séquences et la séquence complémentaire de l'autr ) avec lesdits acides nucléiques, font partie de l'invention, notamment les acides nucléiques variants ou homologues, en particulier les séquences nucléiques de variants alléliques  
15 du gène SPG4 de séquence SEQ ID No. 1 ou de son ADNc de séquence SEQ ID No. 2, ainsi que les séquences génomiques des gènes homologues d'autres mammifères tels que la souris.

Dans la présente description, on entendra désigner par "Spg4" le gène de souris homologue au gène humain SPG4.

20 L'utilisation d'une séquence d'acide nucléique selon l'invention comme sonde ou amorce pour le criblage de banque génomique ou d'ADNc fait bien entendu partie de l'objet de la présente invention.

Sous un autre aspect, l'invention comprend un polypeptide purifié ou isolé codé par un acide nucléique selon l'invention, de préférence à l'exception du peptide de 584  
25 acides aminés dont la séquence est identifiée dans la banque de données GenBank sous le numéro d'accèsion AB029006.

Dans la présente description, on utilisera le terme polypeptide pour désigner également une protéine ou un peptide.

De préférence, la présente invention concerne un polypeptide selon l'invention,  
30 caractérisé en ce qu'il comprend une séquence d'acides aminés choisie parmi le groupe suivant :

- la séquence SEQ ID No. 3, correspondant à la spastin humaine codée par la séquence SEQ ID No. 2 de l'ADNc du gène SPG4 humain ;

- la séquence SEQ ID No. 73, correspondant à un fragment de la spastin murine codée par la séquence SEQ ID No. 72 de l'ADNc incomplet du gène Spg4 de souris, la séquence SEQ ID No. 73 est représentée à la figure 4A, ligne "SPAST\_MOUSE" ;
- la séquence SEQ ID No. 107, correspondant à la spastin murine codée par la séquence SEQ ID No. 106 de l'ADNc complet du gène Spg4 de souris ;
- les séquences de polypeptides homologues et variants du polypeptide de séquence SEQ ID No. 3, SEQ ID No. 73 ou SEQ ID No. 107 ; et
- les séquences de leurs fragments d'au moins 8, 10, 15, 30 ou 50 acides aminés consécutifs.

De manière également préférée, l'invention a pour objet un polypeptide selon l'invention, caractérisé en ce qu'il comprend une séquence d'acides aminés choisie parmi le groupe comprenant :

- a) la séquence SEQ ID No. 3, la séquence SEQ ID No. 73, la séquence SEQ ID No. 107 ou la séquence d'au moins 10 acides aminés consécutifs d'une de ces séquences ; et
- b) les séquences homologues ou variantes des séquences SEQ ID No. 3, SEQ ID No. 73 ou SEQ ID No. 107.

De manière également préférée, l'invention a pour objet un polypeptide selon l'invention, caractérisé en ce qu'il comprend la séquence d'au moins 8, de préférence d'au moins 10, 15, 20 ou 30 acides aminés consécutifs de la séquence du fragment aa 197-228, extrémités incluses, de la séquence SEQ ID No. 3.

De manière également préférée, l'invention a pour objet un polypeptide selon l'invention, caractérisé en ce qu'il comprend une séquence d'acides aminés choisie parmi le groupe suivant :

- la séquence SEQ ID No. 3, la séquence SEQ ID No. 73 et la séquence SEQ ID No. 107, lesquelles séquences portant au moins une des mutations correspondant à un polymorphisme naturel chez l'Homme, notamment celles dont la nature et la localisation sont identifiées au tableau 5 ci-après ou celles qui pourront être identifiées par les méthodes d'identification de mutation du gène SPG4 selon la présente invention ; et
- les séquences de leurs fragments d'au moins 8, 10, 15, 30 ou 50 acides aminés consécutifs.

Il doit être compris que l'invention ne concerne pas les polypeptides sous forme naturelle, c'est-à-dire qu'ils ne sont pas pris dans leur environnement. En effet, l'invention concerne les peptides obtenus par purification à partir de sources naturelles,

ou bien obtenus par recombinaison génétique, ou encore par synthèse chimique et pouvant alors comporter des amino-acides non naturels. La production d'un polypeptide recombinant, qui peut être réalisée en utilisant l'une des séquences nucléotidiques selon l'invention est particulièrement avantageuse car elle permet

5 d'obtenir un niveau de pureté accrue du polypeptide désiré.

Par polypeptide homologue, on entendra désigner les polypeptides présentant certaines modifications par rapport au polypeptide de référence, comme en particulier une ou plusieurs délétions, troncations, un allongement, une fusion chimérique, et/ou une ou plusieurs substitutions, et dont la séquence d'acides aminés présente au moins

10 80 %, de préférence 90 % ou 95 %, d'identité après alignement avec la séquence d'acides aminés de référence.

Par « polypeptide variant » (ou variant protéique), on entendra désigner l'ensemble des polypeptides codés par les séquences nucléiques variantes telles que précédemment définies.

En particulier, les polypeptides variants comprendront tout polypeptide codé par la séquence génomique mutée du gène SPG4 de séquence SEQ ID No. 1, et dont la séquence d'acides aminés présente au moins une mutation correspondant notamment à une troncation, délétion, substitution et/ou addition de résidus d'acides aminés par rapport à la séquence SEQ ID No. 3. Dans le cas présent, les polypeptides variants

15 20 présentant au moins une mutation seront liés aux pathologies de type PSF-AD.

On entendra également désigner par polypeptide variant tout polypeptide résultant de mutation d'un site d'épissage dans la séquence nucléique génomique SEQ ID No. 1.

L'invention comprend également les vecteurs de clonage et/ou d'expression contenant une séquence d'acide nucléique selon l'invention.

25

Les vecteurs selon l'invention, caractérisés en ce qu'ils comportent les éléments permettant l'expression et/ou la sécrétion desdites séquences dans une cellule hôte, ou encore une séquence d'adressage cellulaire, font également partie de l'invention.

Les vecteurs caractérisés en ce qu'ils comportent une séquence de promoteur et/ou de régulateur selon l'invention, font également partie de l'invention.

30

Lesdits vecteurs comporteront de préférence un promoteur, des signaux d'initiation et de terminaison de la traduction, ainsi que des régions appropriées de régulation de la transcription. Ils doivent pouvoir être maintenus de façon stable dans la

cellule et peuvent éventuellement posséder des signaux particuliers spécifiant la sécrétion de la protéine traduite.

Ces différents signaux de contrôle sont choisis en fonction de l'hôte cellulaire utilisé. A cet effet, les séquences d'acide nucléique selon l'invention peuvent être  
5 insérées dans des vecteurs à réplication autonome au sein de l'hôte choisi, ou des vecteurs intégratifs de l'hôte choisi.

Parmi les systèmes à réplication autonome, on utilisera de préférence en fonction de la cellule hôte, des systèmes de type plasmidique ou viral, les vecteurs viraux pouvant notamment être des adénovirus (Perricaudet et al., 1992), des  
10 rétrovirus, des lentivirus, des poxvirus ou des virus herpétiques (Epstein et al., 1992). L'homme du métier connaît les technologies utilisables pour chacun de ces systèmes.

Lorsque l'on souhaitera l'intégration de la séquence dans les chromosomes de la cellule hôte, on pourra utiliser par exemple des systèmes de type plasmidique ou viral ; de tels virus seront, par exemple, les rétrovirus (Temin, 1986), ou les AAV  
15 (Carter, 1993).

Parmi les vecteurs non viraux, on préfère les polynucléotides nus tels que l'ADN nu ou l'ARN nu selon la technique développée par la société VICAL, les chromosomes artificiels de levure (YAC, yeast artificial chromosome) pour l'expression dans la levure, les chromosomes artificiels de souris (MAC, mouse artificial chromosome) pour  
20 l'expression dans les cellules murines et de manière préférée les chromosomes artificiels d'homme (HAC, human artificial chromosome) pour l'expression dans les cellules humaines.

De tels vecteurs seront préparés selon les méthodes couramment utilisées par l'homme du métier, et les clones en résultant peuvent être introduits dans un hôte  
25 approprié par des méthodes standard, telles que par exemple la lipofection, l'électroporation, le choc thermique.

L'invention comprend en outre les cellules hôtes, notamment les cellules eucaryotes et procaryotes, transformées par les vecteurs selon l'invention ainsi que les animaux transgéniques, excepté l'Homme, comprenant une desdites cellules  
30 transformées selon l'invention.

Parmi les cellules utilisables à ces fins, on peut citer bien entendu les cellules bactériennes (Olins et Lee, 1993), mais également les cellules de levure (Buckholz, 1993), de même que les cellules animales, en particulier les cultures de cellules de mammifères (Edwards et Aruffo, 1993), et notamment les cellules d'ovaire de hamster  
35 chinois (CHO), mais également les cellules d'insectes dans lesquelles on peut utiliser

des procédés mettant en œuvre des baculovirus par exemple (Luckow, 1993). Un hôte cellulaire préféré pour l'expression des protéines de l'invention est constitué par les cellules CHO.

5 Parmi les mammifères selon l'invention, on préférera des animaux tels que les souris, les rats ou les lapins, exprimant un polypeptide selon l'invention.

10 Parmi les mammifères selon l'invention, on préférera également ceux comprenant une cellule transformée caractérisée en ce que la séquence de l'un au moins des deux allèles du gène SPG4 contient une au moins des mutations correspondant à un polymorphisme naturel chez l'Homme, notamment celles dont la nature et la localisation sont identifiées au tableau 5 ci-après ou celles qui pourront être identifiées par les méthodes d'identification de mutation du gène SPG4 selon la présente invention.

15 Parmi les mammifères selon l'invention, on préférera également des animaux tels que les souris, les rats ou les lapins, caractérisés en ce que le gène codant pour la spastin selon l'invention, n'est pas fonctionnel ou est invalidé.

Parmi les modèles animaux plus particulièrement intéressants ici, on trouve notamment :

- 20 - les animaux transgéniques présentant au moins dans une de leurs deux séquences alléliques du gène SPG4, une au moins des mutations dont la position et la nature sont identifiées au tableau 5 ou identifiées par une méthode selon la présente invention. Ces animaux transgéniques sont obtenus par exemple par recombinaison homologue sur cellules souches embryonnaires, transfert de ces cellules souches à des embryons, sélection des chimères affectées au niveau des lignées reproductrices, et croissance desdites chimères ;
- 25 - les animaux (de préférence souris) transgéniques surexprimant le gène SPG4 dans lequel pourra être introduite une desdites mutations selon l'invention. Les souris sont obtenues par exemple par transfection de copie de ce gène sous contrôle d'un promoteur fort de nature ubiquitaire, ou sélectif d'un type de tissu, ou après transcription virale ;
- 30 - les animaux (de préférence souris) transgéniques rendus déficients pour le gène SPG4 selon l'invention, par inactivation à l'aide du système LOXP/CRE recombinase (Rohlmann et al., 1996) ou de tout autre système d'inactivation de l'expression de ce gène.

35 Les cellules et mammifères selon l'invention sont utilisables dans une méthode de production d'un polypeptide selon l'invention, comme décrit ci-dessous, et peuvent

également servir à titre de modèle d'analyse et pour le criblage de banque d'ADN (génomique ou d'ADNc).

Les cellules ou mammifères transformés tels que décrits précédemment peuvent être ainsi utilisées à titre de modèles afin d'étudier les interactions entre les polypeptides selon l'invention, et les composés chimiques ou protéiques, impliqués directement ou indirectement dans les activités des polypeptides selon l'invention, ceci afin d'étudier les différents mécanismes et interactions mis en jeu.

Surtout, ils peuvent être utilisés pour la sélection de produits interagissant avec les polypeptides selon l'invention, notamment la spastin humaine de séquence SEQ ID No. 3 ou ses variants selon l'invention, à titre de cofacteur, ou d'inhibiteur, notamment compétitif, ou encore ayant une activité agoniste ou antagoniste de l'activité des polypeptides selon l'invention. De préférence, on utilisera lesdites cellules transformées ou animaux transgéniques à titre de modèle permettant, notamment, la sélection de produits permettant de lutter contre la pathologie liée au gène SPG4 mentionnée ci-dessus.

L'invention concerne également l'utilisation de cellule, de mammifère ou de polypeptide selon l'invention pour le criblage de composé chimique ou biochimique pouvant interagir directement ou indirectement avec les polypeptides selon l'invention, et/ou capable de moduler l'expression ou l'activité de ces polypeptides.

L'invention concerne également l'utilisation d'une séquence d'acide nucléique selon l'invention pour la synthèse de polypeptides recombinants.

La méthode de production d'un polypeptide de l'invention sous forme recombinante est elle-même comprise dans la présente invention, et se caractérise en ce que l'on cultive les cellules transformées, notamment les cellules ou mammifères de la présente invention, dans des conditions permettant l'expression d'un polypeptide recombinant codé par une séquence d'acide nucléique selon l'invention, et que l'on récupère ledit polypeptide recombinant.

Les polypeptides recombinants, caractérisés en ce qu'ils sont susceptibles d'être obtenus par ladite méthode de production, font également partie de l'invention.

Les polypeptides recombinants obtenus comme indiqué ci-dessus, peuvent aussi bien se présenter sous forme glycosylée que non glycosylée et peuvent présenter ou non la structure tertiaire naturelle.

Ces polypeptides peuvent être produits à partir des séquences d'acide nucléique définies ci-dessus, selon les techniques de production de polypeptides recombinants connues de l'homme du métier. Dans ce cas, la séquence d'acide

nucléique utilisée est placée sous le contrôle de signaux permettant son expression dans un hôte cellulaire.

Un système efficace de production d'un polypeptide recombinant nécessite de disposer d'un vecteur et d'une cellule hôte selon l'invention.

5 Ces cellules peuvent être obtenues par l'introduction dans des cellules hôtes d'une séquence nucléotidique insérée dans un vecteur tel que défini ci-dessus, puis la mise en culture desdites cellules dans des conditions permettant la réplication et/ou l'expression de la séquence nucléotidique transfectée.

10 Les procédés de purification de polypeptide recombinant utilisés sont connus de l'homme du métier. Le polypeptide recombinant peut être purifié à partir de lysats et extraits cellulaires, du surnageant du milieu de culture, par des méthodes utilisées individuellement ou en combinaison, telles que le fractionnement, les méthodes de chromatographie, les techniques d'immunoaffinité à l'aide d'anticorps monoclonaux ou polyclonaux spécifiques, etc..

15 Les polypeptides selon la présente invention peuvent être obtenus par synthèse chimique et ce en utilisant l'une des nombreuses synthèses peptidiques connues, par exemple les techniques mettant en oeuvre des phases solides ou des techniques utilisant des phases solides partielles, par condensation de fragments ou par une synthèse en solution classique.

20 La technique de synthèse en phase solide est bien connue de l'homme du métier. Voir notamment Stewart et al. (1984) et Bodansky (1984).

Les polypeptides obtenus par synthèse chimique et pouvant comporter des acides aminés non naturels correspondants sont également compris dans l'invention.

25 Les anticorps mono- ou polyclonaux ou leurs fragments, anticorps chimériques ou immunoconjugués, caractérisés en ce qu'ils sont capables de reconnaître spécifiquement un polypeptide selon l'invention, font partie de l'invention.

30 Des anticorps polyclonaux spécifiques peuvent être obtenus à partir d'un sérum d'un animal immunisé contre les polypeptides selon l'invention, notamment produit par recombinaison génétique ou par synthèse peptidique, selon les modes opératoires usuels.

On notera notamment l'intérêt d'anticorps reconnaissant de façon spécifique certains polypeptides, variants, ou leurs fragments immunogènes, selon l'invention.

Les anticorps monoclonaux spécifiques peuvent être obtenus selon la méthode classique de culture d'hybridomes décrite par Köhler et Milstein, 1975.

Les anticorps selon l'invention sont, par exemple, des anticorps chimériques, des anticorps humanisés, des fragments Fab ou F(ab')<sub>2</sub>. Ils peuvent également se présenter sous forme d'immunoconjugués ou d'anticorps marqués afin d'obtenir un signal détectable et/ou quantifiable.

- 5 L'invention concerne également des méthodes pour la détection et/ou la purification d'un polypeptide selon l'invention, caractérisées en ce qu'elles mettent en œuvre un anticorps selon l'invention.

L'invention comprend en outre des polypeptides purifiés, caractérisés en ce qu'ils sont obtenus par une méthode selon l'invention.

Par ailleurs, outre leur utilisation pour la purification des polypeptides, les anticorps de l'invention, en particulier les anticorps monoclonaux, peuvent également être utilisés pour la détection de ces polypeptides dans un échantillon biologique.

- Ils constituent ainsi un moyen d'analyse immunocytochimique ou immunohistochimique de l'expression des polypeptides selon l'invention, notamment le polypeptide de séquence SEQ ID No. 3 ou l'un de ses variants, sur des coupes de  
15 tissus spécifiques, par exemple par immunofluorescence, marquage à l'or, immunoconjugués enzymatiques.

- Ils pourront permettre notamment de mettre en évidence une expression anormale de ces polypeptides dans les tissus ou prélèvements biologiques, ce qui les rend utiles pour le suivi de l'évolution de la maladie et le diagnostic moléculaire.  
20

Plus généralement, les anticorps de l'invention peuvent être avantageusement mis en œuvre dans toute situation où l'expression d'un polypeptide selon l'invention, normal ou muté, doit être observée.

- Font également partie de l'invention, les méthodes de détermination d'une variabilité allélique, d'une mutation, d'une délétion, d'une perte d'hétérozygotie ou de toute anomalie génétique du gène SPG4 selon l'invention, caractérisées en ce qu'elles mettent en œuvre une séquence d'acide nucléique ou un anticorps selon l'invention.  
25

- La présente invention comprend ainsi une méthode de diagnostic génotypique de la pathologie associée au gène SPG4, caractérisée en ce que l'on met en œuvre une séquence d'acide nucléique selon l'invention.  
30

De préférence, l'invention concerne une méthode de diagnostic génotypique de la maladie associée à la présence d'au moins une mutation sur une séquence du gène SPG4 à partir d'un prélèvement biologique d'un patient, caractérisé en ce qu'il comporte les étapes suivantes :

- a) le cas échéant, isolement de l'ADN génomique à partir de l'échantillon biologique à analyser, ou obtention d'ADNc à partir de l'ARN de l'échantillon biologique ;
- b) amplification spécifique de ladite séquence d'ADN du gène SPG4 susceptible de contenir une mutation à l'aide d'amorces selon l'invention ;
- 5 c) analyse des produits d'amplification obtenus et comparaison de leur séquence avec la séquence normale correspondante du gène SPG4.

L'invention comprend également une méthode de diagnostic de la maladie associée à une expression anormale d'un polypeptide codé par le gène SPG4, notamment le polypeptide de séquence SEQ ID No. 3, caractérisée en ce que l'on met  
10 en contact un ou des anticorps selon l'invention avec le matériel biologique à tester, dans des conditions permettant la formation éventuelle de complexes immunologiques spécifiques entre ledit polypeptide et le ou lesdits anticorps, et en ce que l'on détecte et/ou quantifie les complexes immunologiques éventuellement formés.

Ces méthodes visent par exemple les méthodes de diagnostic de la PSF-AD associée à la présence de mutation dans le gène SPG4 selon l'invention, notamment  
15 anténatal, en déterminant à partir d'un prélèvement biologique du patient la présence de mutations dans au moins une des séquences décrites précédemment. Les séquences d'acides nucléiques analysées pourront aussi bien être de l'ADN génomique, de l'ADNc, ou de l'ARNm.

20 Des acides nucléiques ou anticorps basés sur la présente invention pourront également être utilisés pour permettre un diagnostic positif chez un malade ou un diagnostic pré-symptomatique chez un sujet à risque, notamment avec antécédent familial.

Les méthodes permettant de mettre en évidence une mutation dans un gène  
25 par rapport au gène sauvage sont, bien entendu, très nombreuses. On peut essentiellement les diviser en deux grandes catégories. Le premier type de méthode est celui dans lequel la présence d'une mutation est détectée par comparaison de la séquence mutée avec la séquence correspondante sauvage, et le second type est celui dans lequel la présence de la mutation est détectée de façon indirecte, par  
30 exemple par évidence de mésappariements dus à la présence de la mutation.

Ces méthodes peuvent mettre en oeuvre les sondes et amorces de la présente invention décrites. Il s'agit généralement de séquences nucléiques d'hybridation purifiées comprenant au moins 15 nucléotides, de préférence 20, 25 ou 30 nucléotides, caractérisées en ce qu'elles peuvent s'hybrider spécifiquement avec une séquence  
35 nucléique selon l'invention.

De préférence, les conditions d'hybridation spécifiques sont telles que celles définies précédemment ou dans les exemples. La longueur de ces séquences nucléiques d'hybridation peut varier de 15, 20 ou 30 à 200 nucléotides, particulièrement de 20 à 50 nucléotides.

5            Parmi les méthodes de détermination d'une variabilité allélique, d'une mutation, d'une délétion, d'une perte d'hétérozygotie ou d'une anomalie génétique, on préfère les méthodes comprenant au moins une étape d'amplification dite par PCR (réaction en chaîne par la polymérase) ou par PCR-like de la séquence cible selon l'invention susceptible de présenter une anomalie à l'aide de couple d'amorces de séquences  
10            nucléotidiques selon l'invention. Les produits amplifiés pourront être traités à l'aide d'enzyme de restriction appropriée avant de procéder à la détection ou au dosage du produit ciblé.

             Les mutations du gène SPG4 selon l'invention, peuvent être responsables de différentes modifications de son produit de traduction, modifications utilisables pour  
15            une approche diagnostique. En effet, les modifications d'antigénicité liées à ces mutations peuvent permettre la mise au point d'anticorps spécifiques. La discrimination du produit de gène muté peut être réalisée par ces méthodes. Toutes ces modifications peuvent être utilisées dans une approche diagnostique grâce à plusieurs méthodes bien connues basées sur l'utilisation d'anticorps mono- ou polyclonaux reconnaissant  
20            le polypeptide normal ou des variants mutés, comme par exemple par RIA ou par ELISA.

             Ainsi, l'invention a également pour objet une trousse ou nécessaire pour le diagnostic, notamment pour le diagnostic de la PSF-AD associée à la présence de mutation dans le gène SPG4 selon l'invention, caractérisée en ce qu'elle comprend au  
25            moins un composé choisi parmi le groupe de composés suivant :

- a) un acide nucléique, notamment comme amorce ou sonde, selon la présente invention ; et
- b) un anticorps selon l'invention.

             Sous un autre aspect, l'invention comprend une méthode de sélection d'un  
30            composé chimique ou biochimique capable de prévenir et/ou de traiter la PSF-AD associée au gène SPG4, caractérisée en ce que l'on met en œuvre une séquence d'acide nucléique selon l'invention, un polypeptide selon l'invention, un vecteur selon l'invention, une cellule selon l'invention, un mammifère selon l'invention ou un anticorps selon l'invention.

Sont également comprises dans l'invention, les méthodes de sélection de composés chimiques ou biochimiques capables d'interagir directement ou indirectement avec des polypeptides selon l'invention ou avec les acides nucléiques selon l'invention, et/ou permettant de moduler l'expression ou l'activité de ces polypeptides, caractérisées en ce qu'elles comprennent la mise en contact d'un polypeptide selon l'invention, d'une cellule transformée selon l'invention, ou d'un mammifère selon l'invention, avec un composé candidat et, la détection d'une modification de l'activité dudit polypeptide.

Par exemple, mais sans s'y limiter, on peut citer une méthode d'identification de molécules capables d'interagir avec un polypeptide selon l'invention en utilisant un système de double hybride bactérien ou levure tel que le Matchmaker Two Hybrid System 2, selon les instructions du manuel accompagnant le Matchmaker Two Hybrid System 2 (Catalogue N° K1604-1, Clontech).

Les acides nucléiques codant pour des protéines interagissant avec les séquences promotrices et/ou régulatrices du gène SPG4 selon l'invention, peuvent être criblés et/ou sélectionnés en utilisant un système de simple hybride tel que celui décrit dans le manuel accompagnant le kit Matchmaker One-Hybrid System de Clontech (Catalog N° K1603-).

Sous un autre aspect, l'invention comprend l'utilisation d'acide nucléique ou de polypeptide selon l'invention, d'un vecteur selon l'invention, d'une cellule selon l'invention, ou d'un mammifère selon l'invention, pour l'étude de l'expression ou de l'activité du gène SPG4.

D'autres caractéristiques et avantages de l'invention apparaissent dans la suite de la description avec les exemples et les figures dont les légendes sont représentées ci-après.

#### LEGENDES DES FIGURES

FIGURES 1A, 1B et 1C : Carte physique de l'intervalle SPG4 et organisation génomique de SPG4.

FIGURE 1A : La région candidate de 1,5 Mb est délimitée par les marqueurs génétiques D2S352 et D2S2347 indiqués en caractères gras. La position des marqueurs polymorphes et autres STSs est indiquée en caractères standards alors que la position des ESTs est indiquée en italique. Les clones de BAC constituant la carte de préséquénage sont représentés par des rectangles dont le nom figure au-

dessus et la taille précise du clone au-dessous si celle-ci a pu être déterminée. Le nom des BACs A, B, C, ... est suivi d'une parenthèse contenant le nom du clone précédé d'un « b » si le clone est issu de la banque de BACs CITB\_978\_SKB ou d'un « B » si celui-ci provient de la banque RPCI-11.

- 5           FIGURE 1B : Représentation schématique du gène SPG4 qui chevauche les BACs D (b336P14) et G (B563N4). Les exons sont figurés comme des rectangles noirs avec leur nom au-dessus.

- FIGURE 1C : Les cinq mutations identifiées dans sept familles de PSF-AD liées au locus SPG4 sont positionnées dans les exons 7, 11, 13 et dans le site accepteur d'épissage de l'intron 15.
- 10

FIGURE 2 : Séquence nucléique et protéique de l'ADNc SPG4 et de la spastin.

- Les 17 barres verticales avec un nombre situé au-dessous représentent les jonctions entre les différents exons. Le codon ATG initiateur est localisé en position nt 126-128 et le codon STOP de terminaison en position nt 1974-1976. Cinq des mutations identifiées à ce jour, dont la perte de l'exon 16, sont indiquées en italique (nt 1210, nt 1468, nt 1520, nt 1620 et pour la perte de l'exon 16 : nt 1813-1853). Le site de polyadénylation est en italique et souligné. Le signal de localisation nucléaire (NLS) putatif, RGKKK, ainsi que les trois domaines conservés prédits par l'analyse dans la base de données ProDom sont respectivement localisés aux positions aa7-11 (NLS),
- 15
- 20           aa342-409 (domaine 92), aa411-509 (domaine 179) et aa512-599 (domaine 6226). Les quatre motifs prédits par la comparaison de séquence dans la base de données Prosite sont : deux motifs «leucine zippers» aux positions aa50-78 et aa508-529, le site de fixation de l'ATP (ou motif A de Walker) aux positions aa382-389 et le domaine de dimérisation « helix-loop-helix » aux positions aa478-486. Les motifs A et B de
- 25           Walker, «GPPGNGKT» et «IIFIDE», ainsi que le consensus minimal des AAA sont soulignés.

FIGURES 3A et 3B : Caractérisation d'une mutation d'un site d'épissage chez les individus atteints de trois familles de PSF-AD liée au locus SPG4.

- FIGURE 3A : Amplification par PCR du fragment IV de l'ADNc SPG4 à partir d'ADNc de lymphoblastes : puits M, marqueur de taille VII (Boehringer) ; puits 1, membre non atteint de la famille 2992 ; puits 2, patient de la famille 2992 ; puits 3, membre non atteint de la famille 5330 ; puits 4, patient de la famille 5330 ; puits 5, patient de la famille 5226 ; puits 6, témoin négatif (ADN génomique humain).
- 30

- FIGURE 3B : Graphe de séquence de la mutation du site accepteur d'épissage de l'intron 15.
- 35

Séquence génomique de l'individu contrôle en haut et d'un patient de la famille 2992 en bas. L'astérisque à la position nt 1813-4 indique un polymorphisme A->C qui touche un nucléotide non conservé du site accepteur d'épissage de l'intron 15 chez le patient.

5 FIGURES 4A et 4B : Les homologues de la spastin.

Les résidus identiques sont respectivement surlignés en grisé.

FIGURE 4A : Alignement multiple créé par CLUSTAL W de huit protéines issues de divers organismes et présentant une forte homologie de séquence avec les spastin humaine et murine (SEQ ID No. 73).

10 FIGURE 4B : Alignement par CLUSTAL W des métalloprotéases de levure AFG3, RCA1 et YME1, et des paraplegin et spastin humaines.

FIGURE 5 : Alignement par BLASTN des séquences nucléiques de l'ADNc SPG4 et de son orthologue de souris Spg4 (SEQ ID No. 72). Le site de polyadénylation de l'ADNc murin est souligné et en italique. Le codon STOP est localisé en position nt 1515-1517 dans l'ADNc murin et en position nt 1974-1976 dans l'ADNc humain.

15 FIGURES 6A, 6B et 6C : Analyse par PCR de l'expression de SPG4 et de son orthologue murin Spg4.

FIGURE 6A : Collection d'ADNc provenant de multiples tissus de souris.

20 Puits M, marqueur de taille V (Boehringer) ; puits 1, cœur ; puits 2, cerveau ; puits 3, rate ; puits 4, poumon ; puits 5, foie ; puits 6, muscle squelettique ; puits 7, rein ; puits 8, testicule ; puits 9, embryon de 7 jours E7 ; puits 10, embryon de 11 jours E11 ; puits 11, embryon de 15 jours E15 ; puits 12, embryon de 17 jours E17 ; puits 13, témoin négatif (ADN génomique de souris).

FIGURE 6B : Collection d'ADNc provenant de multiples tissus humains.

25 Puits M, marqueur de taille VII (Boehringer) ; puits 1, cerveau ; puits 2, cœur ; puits 3, rein ; puits 4, foie ; puits 5, poumon ; puits 6, pancréas ; puits 7, placenta ; puits 8, muscle squelettique ; puits 9, témoin négatif (ADN génomique humain) ; puits 10, témoin négatif (pas d'ADN).

30 FIGURE 6C : Collection d'ADNc provenant de multiples tissus de fœtus humain.

Puits M, marqueur de taille VII (Boehringer) ; puits 1, cerveau ; puits 2, cœur ; puits 3, rein ; puits 4, foie ; puits 5, poumon ; puits 6, muscle squelettique ; puits 7, rate ; puits 8, thymus ; puits 9, témoin négatif (ADN génomique humain) ; puits 10, témoin négatif (pas d'ADN).

## EXEMPLES

### Exemple 1 : Matériels et méthodes

#### 1) Sous-clonage et séquençage de la région candidate

Douze BACs provenant de deux banques génomiques humaines, CITB\_978\_SKB (commercialisées par Research Genetics) et RPCI-11 (Osoegawa et al., 1998), et recouvrant l'intervalle SPG4 ont été sélectionnés pour être séquencés (Hazan et al., Genomics, 60 (3), 309-19, 1999). 40 µg de l'ADN de chaque BAC a été digéré partiellement par l'enzyme de restriction CviJI (CHIMERx) et séparé par électrophorèse sur gel d'agarose 0,4 % LMP (FMC). Des fractions d'ADN dont les tailles varient autour de 3, 5 et 10 kb ont été éluées avec la β-agarase (Biolabs) et liguées à un vecteur plasmide pBAM3 préalablement digéré par SmaI et déphosphorylé dans un rapport de 1Xinsert pour 5Xvecteur. Des bactéries E. coli DH10B électrocompétentes (GIBCO-BRL) ont été transformées par électroporation avec les différentes ligations. Environ 1000 à 1500 sous-clones par BAC (8 à 10 génomes équivalents) composés de 20 % de clones avec inserts à 10 kb, 40 % de clones avec inserts à 5 kb et 40 % de clones avec inserts à 3 kb ont été isolés. Les extrémités des inserts de ces clones ont été séquencées sur un séquenceur automatique LICOR 4200. Pour chaque BAC, les séquences furent assemblées en un squelette constitué de plusieurs contigs à l'aide des logiciels Phred et Phrap. Les trous entre chaque contig furent séquencés avec des dideoxynucléotides marqués sur un séquenceur ABI 377 (PE-Applied Biosystems). Les exons contenus dans ces contigs de séquence ont été prédits par les programmes informatiques GRAIL II, GENSCAN, FGENEH et Genie. Les séquences furent également comparées dans les bases de données nucléiques et protéiques de l'EMBL et de GenBank avec les programmes BLASTN et BLASTX. La détermination des séquences promotrices fut réalisée par les programmes informatiques TSSG et TSSW. Les résultats de toutes ces analyses d séquences furent visualisés par le programme d'annotation de séquence Genotator.

#### 2) Clonage de l'ADNc

L'ADNc du gène SPG4 a été isolé par des expériences de RACE-PCR en 5' et 3' sur des ARN polyA<sup>+</sup> de cerveau fœtal, cerveau adulte et foie adulte à l'aide du kit d'amplification d'ADNc Marathon (Clontech), selon les instructions du fournisseur. Une première PCR, suivie d'une PCR interne furent effectuées avec différents couples d'amorces dont les séquences sont indiquées sur le tableau 1 ci-après :

**Tabl au 1**  
**Amorces utilisées pour les RACE-PCR et les amplifications d'ADNc**

Amorce	Séquence (5'-3')	Position en 5' couple/	PCR	Taille du produit
SPA_5RACE5	CGGAGCTCCTCTTGGCTGCCATG (SEQ ID No.4)	nt 405		
SPA_5RACE6	AGAAGCGCTGGCAGAGCCACACGAAG (SEQ ID No.5)	nt 372		
SPA_5RACE7	AAGGCGACCAAACGCAGCAGCGCGAAG (SEQ ID No.6)	nt 331		
SPA_3RACE1	AGGAGCAAGCTGTGGAATGGTATAAG (SEQ ID No.7)	nt 550		
SPA_3RACE2	TGGTTATGGCCAAGGACCGCTTACAAC (SEQ ID No.8)	nt 689		
SPA_3RACE3	CAAACGGACGTCTATAATGACAGTAC (SEQ ID No.9)	nt 747		
SPA_3RACE4	TTAGGAATGTGGACAGCAACCTTGC (SEQ ID No.10)	nt 1075		
SPA_3RACE5	CTTCTCTGAGGCCTGAGTTGTTTAC (SEQ ID No.11)	nt 1207		
SPA_3RACE6	TGCTAGAATGACTGATGGATACTCAGG (SEQ ID No.12)	nt 1736		
SPA_3RACE7	AGATGCAGCACTGGGTCCTATCCG (SEQ ID No.13)	nt 1787		
SPA_3RACE8	ATGAACGTCATCGGCTACAGAAACAG (SEQ ID No.14)	nt 2037		
SPA_Db	TAGCAGTGGCTGCCGCCGT (SEQ ID No.15)	nt 45	b+m	655 pb
SPA_Dm	AAGCGGTCCTTGGCCATAAC (SEQ ID No.16)	nt 700		
SPA_Dc	GGCGGCAGTGAGAGCTGTG (SEQ ID No.17)	nt 106	c+n	543 pb
SPA_Dn	CTAGCTCTTTCACACTGTTC (SEQ ID No.18)	nt 649		
SPA_Ad	AACAGGCCTTCGAGTACATC (SEQ ID No.19)	nt 487	d+n	746 pb
SPA_Am	CTGTGAACAACTCAGGCCTC (SEQ ID No.20)	nt 1233		
SPA_Ac	ATGAGAAAGCAGGACAGAAG (SEQ ID No.21)	nt 532		
SPA_An	TGCCAAGTCTTGACCAGC (SEQ ID No.22)	nt 1175		
SPA_Ba	CTACAACTGCTACTCGTAAG (SEQ ID No.23)	nt 1036	a+m	763 pb
SPA_Bm	CAGTGCTGCATCTTTTGCC (SEQ ID No.24)	nt 1799		
SPA_Bb	TAGGAATGTGGACAGCAACC (SEQ ID No.25)	nt 1076		
SPA_Bn	AAAGCTGTTAGGTCACCTCC (SEQ ID No.26)	nt 1780		
SPA_Ca	TGGAGATGACAGAGTACTTG (SEQ ID No.27)	nt 1550	a+m	766 pb
SPA_Cm	CTGGAATACTTTCATCTGC (SEQ ID No.28)	nt 2316		
SPA_Cb	ATGAGGCTGTTCTCAGGCG (SEQ ID No.29)	nt 1603		

Les produits de RACE-PCR ont été clonés avec le kit TA-cloning (Invitrogen) et les clones correspondants ont été séquencés sur un ABI 377 (PE-Applied Biosystems). La séquence du transcrit SPG4 a été vérifiée par le séquençage de produits de PCR amplifiés à partir d'une population d'ADNc provenant des lymphoblastes de 6 individus sains.

### 3) Détection des mutations

Les ARNs totaux ont été extraits de lignées de lymphoblastes d'un individu atteint par famille étudiée et de 6 individus témoins à l'aide du kit RNA PLUSR (bioprobe System). La synthèse de l'ADNc a été réalisée sur 500 ng à 1 µg d'ARN avec 100 pmoles d'amorces hexamères aléatoires (Pharmacia) et 200 unités de reverse transcriptase Superscript II (Gibco BRL) dans des conditions standards. Quatre amplifications par PCR, générant des fragments chevauchants qui recouvrent la totalité de la phase ouverte de lecture de SPG4, ont été réalisées sur les ADNc des patients et contrôles. Le fragment I fut amplifié avec les amorces SPA\_Db/SPA\_Dm, puis en PCR interne avec les amorces SPA\_Dc/SPA\_Dn. Les fragments II, III, et IV ont été respectivement amplifiés avec les amorces SPA\_Ad/SPA\_Am, SPA\_Ba/SPA\_Bm et SPA\_Ca/SPA\_Cm (cf. les séquences de ces amorces sur le tableau 1). Chaque amplification a été effectuée dans un volume total de 50 µl contenant 4 µl d'ADNc (~1/7 ème de la prép.), 20 pmoles de chaque amorce, 200 µM de dNTPs, 50 mM de KCl, 10 mM de Tris pH9, 1,5 mM MgCl<sub>2</sub>, 0,1 % de triton X-100, 0,01 % de gélatine et 2,5 unités de Taq polymérase (Cetus-PE). Les réactions de PCR ont été réalisées selon le procédé du « hot start » : la Taq polymérase est ajoutée à 92°C après une première étape de dénaturation de 5 min à 94°C. Les échantillons sont par la suite soumis à 35 cycles de dénaturation (94°C pendant 40 sec), d'hybridation (55°C pendant 50 sec, à l'exception du fragment I : 58°C pendant 50 sec) et d'élongation (72°C pendant 1 min), suivis d'une dernière étape d'élongation (5 min à 72°C). Les produits de PCR sont séquencés sur un séquenceur automatique ABI 377 (PE-Applied Biosystems) avec les amorces SPA\_Dc/SPA\_Dn, SPA\_Ac/SPA\_An, SPA\_Bb/SPA\_Bn et SPA\_Cb/SPA\_Cm pour les fragments I, II, III et IV respectivement.

Les mutations furent également recherchées ou confirmées par séquençage des 17 exons prédits du gène SPG4 chez les patients et contrôles. Chaque exon fut amplifié avec le couple correspondant d'amorces « a+m » (cf. tableau 2 ci-après), à l'exception de l'exon 1 (gSPAex1c/gSPAex1m), et des exons 10, 11 et 12 qui furent co-amplifiés avec les couples d'amorces gSPAex10a/gSPAex12m et gSPAex11a/gSPAex12m.

**Tableau 2**  
**Amorces de PCR pour l'amplification et le séquençage des exons**

Exon	Taille du produit	Programme de PCR	Amorce	Séquence (5'-3') (SEQ id Nos. ; 30 à 71)
1	1048 pb	0	gSPAex1c	GTGAGCCGAACTGCACATTG
			gSPAex1m	CAAAGTCGACAGCTACAGTGC
			gSPAex1d	GGAAGTGTAGTTGAGTGGGA
			gSPAex1n	AGATGAGGCTCCGACCTAC
2	624 pb	3	gSPAex2a	AATGCCACACTTGTAATCTC
			gSPAex2m	TGTGAATATATCATAATTTGGG
			gSPAex2b	TACAGCAGTTCTCATGATG
3	812 pb	1	gSPAex3a	GACCAAATTGGTGCATGCATG
			gSPAex3m	ACATTTCCAATACATCCCAC
4	379 pb	3	gSPAex4a	ATTTGTCATTTACATGCAC
			gSPAex4m	TTAGAATGACTATACCTGAC
			gSPAex4n	TCAGGTAAAGTAAGACTC
5	830 pb	4	gSPAex5a	TTCCTATCTACCTAGTGAC
			gSPAex5m	TTTTATAGCAAGTTGCCCTG
			gSPAex5b	CCTATGAAGATCCTGGTAC
6	484 pb	3	gSPAex6a	TGTCATGATTCTAACAAGGG
			gSPAex6m	TCTATTTCACTCCTGACATG
7	420 pb	2	gSPAex7a	GTCATAGGGCTTAGGCTTC
			gSPAex7m	ATCATACTACCCACTTTTCC
8	647 pb	3	gSPAex8a	TGTTTGGGAAGATGCTACTG
			gSPAex8m	CTACTGAAGATAACGTACATG
9	1268 pb	1	gSPAex9a	CATTGATTGCCATGTATTGG
			gSPAex9m	AGAAGGCCAGAAATACTCAG
			gSPAex9b	GTACTTAAATCGGTAAATATGG
10]	1061 pb	4	gSPAex10a	CTCAAGTCTTAGGAATGCAG
11			gSPAex10b	GCACTTAACCAGGCTGTATG
12]	551 pb	3	gSPAex11a	CTCAGATGACTCACATAGC
			gSPAex12m	CTTTACTAGACTAATTCTCCTG

13	1361 pb	4	gSPAex13a	CAGATTCAAGAAGACAGATC
			gSPAex13m	GCAATAATTCACCACTTG
			gSPAex13n	GGTAGTTCTTGTTTCTGCTC
14	985 pb	4	gSPAex14a	CAAGTGTGGTGAATTATTGC
			gSPAex14m	GAGCTGAAAAGTATTCAGC
			gSPAex14n	TGCAAAGGACATAGCCAGTG
15	1076 pb	1	gSPAex15a	AGCCTCTGGAGATAGTATGC
			gSPAex15m	CTAGAACAGGGGTCACAGTC
			gSPAex15n	TTGGACTTCTTAACTTC
16	1404 pb	4	gSPAex16a	GCAGTATGCAAGAAATTGAAC
			gSPAex16m	GGCCTGTAATTTTCTTCTG
			gSPAex16b	GTACTGAATAGATACATGTAG
17	445 pb	3	gSPAex17a	GTGTAGCAGATCAACATAG
			gSPAex17m	CATCTTCAAGTTTGGTGAC

Hormis l'exon 1, amplifié à l'aide du kit Advantage GC genomic PCR kit (Clontech) selon les instructions du fournisseur, quatre programmes de PCR légèrement différents (1, 2, 3 et 4) furent utilisés pour amplifier les exons de SPG4 (voir tableau 2). Les amplifications furent toutes effectuées dans un volume de 50 µl contenant 100 ng d'ADN génomique, 50 pmoles de chaque amorce, 250 µM de dNTPs, 1X de tampon Takara et 1 unité de Taq polymérase Takara La Taq (Shuzo Co.). Les réactions de PCR ont été réalisées selon le procédé du « hot start » : la Taq polymérase est ajoutée à 94°C après une première étape de dénaturation de 5 min à 96°C. Les échantillons sont par la suite soumis à 30 cycles de dénaturation (94°C pendant 40 sec), d'hybridation (prog. 1 : 60°C pendant 50 sec ; prog. 2, 58°C pendant 50 sec, prog. 3 et 4 : 55°C pendant 50 sec) et d'élongation (prog. 1 et 4 : 72°C pendant 1 min, prog. 2 et 3 : 72°C pendant 40 sec), suivis d'une dernière étape d'élongation (10 min à 72°C). Le séquençage de ces produits de PCR a été réalisé sur un séquenceur ABI 377 (PE-Applied Biosystems) en utilisant soit les amorces de PCR soit les amorces internes notées « b » et « n » (voir tableau 2).

#### 4) Caractérisation de SPG4

Les clones d'ADNc 977312 (EST AA560327) et 568234 (EST AA107866) issus des banques d'ADNc de blastocyste et d'embryon E8 de souris, qui tous deux

correspondent à l'orthologue murin de SPG4, ont été isolés par le consortium IMAGE et séquencés au laboratoire sur un séquenceur ABI 377 (PE-Applied Biosystems). Afin d'analyser le profil d'expression de SPG4 et de son orthologue murin Spg4, les collections d'ADNc de différents tissus humains fœtaux et adultes, ainsi que de tissus de souris (panels MTC, Clontech) ont été testées par PCR selon le protocole du fournisseur avec le couple d'amorces SPA\_Ca/SPA\_Cm pour les ADNc humains et le couple SPA\_Ca /spam (spam : 5'-ACCGAAGTCAAGAGCCTATC-3') pour les ADNc de souris. Les conditions de PCR sont celles utilisées pour l'amplification de SPG4 à partir d'ADNc de lignées de lymphoblastes (cf. § Détection des mutations), excepté que les échantillons ont été soumis à 32 cycles pour les ADNc issus de tissus adultes humains et de tissus murins, et à 28 cycles pour les ADNc issus de tissus fœtaux. Les produits d'amplification ont migré par électrophorèse sur des gels d'agarose 2 %.

5) Analyse histologique d'une biopsie musculaire d'un patient

Les analyses histologiques et histo-enzymatiques ont été réalisées à partir d'une biopsie musculaire d'un patient issu d'une famille liée au locus SPG4 selon les techniques standards décrites dans Casari et al., 1998.

6) Numéros d'accès dans les bases de données publiques

L'ADNc SPG4 (ou SPAST) et la séquence protéique déduite, GenBank/EMBL AJ246001 ; le clone d'ADNc incomplet Spg4, GenBank/EMBL AJ246002 ; le gène SPG4 (ou SPAST), GenBank/EMBL AJ246003.

Exemple 2 : Analyse de la séquence de l'intervalle SPG4

L'analyse des événements de recombinaison a permis de réduire la région candidate SPG4 à un intervalle génétique de 0 cM entre les marqueurs D2S352 et D2S2347 (19, 20). Une carte de préséquencage de l'intervalle SPG4 composée de 37 BACs a été construite (Hazan et al., sous presse dans Genomics) ; la région candidate couvre une distance physique d'environ 1,5 Mb. Douze BACs chevauchants, s'étendant sur l'intervalle SPG4 à l'exception d'un unique trou de 4 kb entre les clones A et E, ont été sélectionnés pour être séquencés (Fig. 1A). Sept de ces BACs (A, B, C, D, E, F et G), couvrant approximativement 70 % de la région d'intérêt, ont déjà été séquencés. Les séquences de ces 7 BACs ont été comparées à celles des bases de données nucléiques et protéiques, et analysées avec quatre programmes de prédiction d'exons. Ces analyses de séquences préliminaires ont permis de mettre en évidence 14 unités de transcription potentielles, dont trois correspondant aux gènes codant pour la xanthine deshydrogénase, la stéroïde 5 $\alpha$ -réductase 2 et une protéine liant le TGF $\beta$ . Sur les 14 gènes détectés par l'analyse de séquence, 9 avaient été préalablement

identifiés dans les bases de données d'EST (pour « Expressed Sequence Tag ») et localisés au sein de l'intervalle SPG4 (Hazan et al., sous presse dans Genomics) ; les 5 gènes restants n'ont pu être identifiés qu'en séquençant la région candidate. L'un de ces 5 nouveaux gènes présentait une homologie en 3' de sa région codante avec les gènes codant pour la famille protéique des AAA (Confalonieri et al., 1995). Des analyses de séquence plus approfondies ont montré que ce gène, nommé SPG4 (ou SPAST), était composé de 17 exons et s'étendait sur une région d'environ 90 kb, couverte par deux clones de BAC adjacents, D et G (cf. Fig. 1B). Les trois premiers exons prédits de ce gène furent identifiés dans le BAC D par deux des quatre programmes de prédiction d'exons utilisés, GRAIL II et GENSCAN ; ils présentent une forte homologie avec un EST de blastocyste de souris, AA560327. Les 14 derniers exons se trouvent dans le BAC G. La séquence protéique déduite des exons 7 à 17 est significativement homologue à une sous-classe de la famille des AAA, comportant les protéines de levure Yta6p (Schnall et al., 1994), TBP6 (Schnall et al., 1994) et End 13, ainsi que la protéine de souris SKD1 (Perier et al., 1994).

Sur les quatre programmes de prédiction d'exons, FGENEH semble le plus fiable et le plus puissant, permettant la détection de la plupart des gènes de cette région chromosomique en 2p21-p22. Cette constatation s'applique également au gène SPG4 pour lequel 15 exons ont pu être mis en évidence par ce programme quand seuls 4, 9 ou 11 exons ont pu respectivement être localisés par les programmes Genie, GRAIL II et GENSCAN. L'organisation génomique de ce gène (Fig. 1B) a pu être confirmée par la suite grâce à la détermination de la séquence de l'ADNc SPG4. Les jonctions introns/exons sont représentées sur le tableau 3 ci-après : la taille des exons varie de 41 pb (exon 16) à 1,410 kb (exon 17), celle des introns variant de 140 pb (intron 11) à 23,247 kb (intron 1).

Tableau 3  
Organisation intron/exon du gène SPG4

Exon/ intron	Taille de l'exon (pb)	Position sur l'ADNc	Site accepteur d'épissage (SEQ ID No. 74 à 89)	Site donneur d'épissage (SEQ ID Nos. 90 à 105)	Taille de l'intron (pb)
1	540	1		TGAGAAAG/gtaactaggggctg9	23 247
2	87	541	atthttattttaaag/CAGGACAG	AGGACAAAG/gtaagattgtattgt	1 943
3	84	628	aatttttcttctcag/GTGAACAG	ACTTCTAG/gtatcaattaatgtat	9 190
4	96	712	cttctgtgtgcatag/AGAAAGATG	CCAGTCAG/gtgggttaggtaac	15 745
5	188	808	acttttccctgtcag/AAAGTGGA	CTCATAAG/gtatctgggacagta	876
6	134	996	tttgtatcccttaag/GGTACTCC	GTGGACAA/gtaagtttgccatct	283
7	94	1 130	aggctgtgttcttag/TGGAACAG	GGCCTGAG/gtaagaaacttatatt	10 735
8	75	1 224	agtatatatttttag/TTGTTTCC	CAATGCTG/gtaaggggtctctca	1 385
9	72	1 299	cttgatatttttaaag/GCTAAAGC	CAAAATAC/gtgagtgctctgttc	8 083
10	76	1 371	taatgccttgttttag/GTGGGAGA	TTTTATAG/gtaagaaacatatctt	238
11	92	1 447	ctgtatttccctctag/ATGAAGTT	TTGATGGT/gtaagtggtgattatg	140
12	80	1 539	gatttttgctgttag/GTACAGTC	GTTCTCAG/gtagggagatttatat	4 715
13	43	1 619	ggatttttttttag/GCGTTTCA	ATGAGGAG/gtatgtatctgtgttt	1 389
14	80	1 662	ttttaataattttcag/ACAAGACT	CTTGCTAG/gtgagtaattggatt	1 521
15	71	1 742	tccctccctcctcag/AATGACTG	TATCCGAG/gtaggtatacaagagc	2 210
16	41	1 813	ctttatgtttttacag/AACTAAAA	CCAGTGAG/gtatagtaatttaaca	7 115
17	1 410	1 854	cttttaaaaatctag/ATGAGAAA		

Les séquences des exons et des introns sont respectivement indiquées en majuscules et minuscules.

### Exemple 3 : Identification de l'ADNc SPG4

Plusieurs amplifications successives par RACE-PCR en 5' et 3' furent réalisées sur des collections d'ADNc de cerveau et de foie adultes et de cerveau fœtal, afin de caractériser le transcrit SPG4. Toutes les RACE-PCR en 5' ont donné des produits

5 d'amplification se terminant à la position nt 263 de l'ADNc SPG4 (Fig. 2), ce qui était probablement dû au contenu riche en GC de la région 5' du transcrit (90 % de GC dans les 60 pb précédant la position nt 263). Quatre produits de PCR chevauchants, recouvrant la totalité de la région codante, ont été amplifiés à partir des ADNc issus des lymphoblastes de six individus contrôles et séquencés intégralement dans le but de

10 vérifier la séquence du transcrit SPG4. L'alignement des séquences de tous les produits de PCR et RACE-PCR a permis de reconstituer une séquence de 3263 pb comprenant une phase ouverte de lecture de 1848 pb précédée par une région 5' non traduite (5' UTR pour « 5' UnTranslated Region ») de 125 pb et suivie par une région 3' UTR de 1290 pb incluant un site de polyadénylation entre les positions nt 3227-3232, ~ 35 pb en amont de

15 la queue polyA (Fig. 2). La comparaison de la séquence de l'ADNc SPG4 avec les banques de données d'ESTs a permis de détecter une homologie significative avec 6 ESTs humains dont l'EST N47973 qui contient une région 3' non codante plus étendue (+ 180 pb) comprenant un deuxième site de polyadénylation. Le site d'initiation de la traduction a été identifié par la présence d'une séquence consensus de Kosak

20 (CTGTGAatgA) définie comme un «contexte adéquat» à l'initiation de la traduction attendu qu'une purine est localisée 3 nt en amont de l'ATG initiateur, lui-même précédé d'un codon STOP. La séquence de l'ADNc de 3263 pb est identique à la séquence transcrite déduite des 17 exons du gène SPG4. L'analyse de la séquence de la région 5' à l'aide des programmes informatiques TSSG et TSSW suggère la présence d'une

25 séquence promotrice de type TATA box située 43 pb en amont de la position nt 1 de l'exon 1.

### Exemple 4 : Mutations dans le gène SPG4

Des mutations hétérozygotes ont été recherchées dans l'ADNc SPG4 provenant de lymphoblastes de 14 patients issus de familles liées au locus SPG4 (1 individu atteint

30 par famille). Quatre fragments de PCR chevauchants I, II, III et IV recouvrant la phase ouverte de lecture de l'ADNc SPG4 ont été amplifiés et séquencés chez les 14 patients ainsi que chez 6 individus sains contrôles. L'électrophorèse sur gel agarose du fragment de PCR IV a montré trois bandes d'intensité égale chez 3 patients des familles 2992, 5226 et 5330 provenant de la même région de Suisse, ce qui suggérerait une microdélétion

35 ou une mutation d'un site d'épissage ; les deux bandes supplémentaires n'étaient pas

présentes chez 2 individus sains issus des familles 2992 et 5330 (Fig. 3A). La séquence génomique de l'exon 16 a révélé une mutation hétérozygote A->G du site accepteur d'épissage (AG) de l'intron 15 chez les individus atteints de ces trois familles (Fig. 3B) ; cette mutation engendre la perte de l'exon 16 suivie d'un décalage de la phase de lecture dans le transcrit anormal. Aucun des membres sains incluant maris et épouses ne porte cette mutation du site d'épissage. L'identification de la même mutation chez tous les membres atteints de ces trois familles suisses démontre l'existence d'un ancêtre commun, ce qui avait été préalablement suggéré par l'étude des haplotypes.

Trois mutations ponctuelles 1210C->G, 1468G->A et 1620C->T qui introduisent des substitutions d'un acide aminé dans la séquence protéique (S362C, C448Y et R499C) ont été respectivement mises en évidence par le séquençage des fragments de PCR III et IV chez les individus atteints des familles 624, 4014 et 618. Ces trois substitutions impliquent toutes un résidu cystéine, induisant la perte ou l'insertion d'une cystéine dans la séquence protéique. Une délétion de 1 pb, 1520delT, qui crée l'apparition d'un codon STOP induisant une protéine tronquée composée de 465 acides aminés (aa) a été détectée chez les individus atteints de la famille A. Aucune des cinq mutations résumées dans le tableau 4 ci-après n'a été trouvée chez les individus contrôles testés, qu'ils appartiennent à la fratrie saine ou aux conjoints des sept familles analysées ici. Ces cinq mutations affectent de façon importante la séquence protéique dans un domaine très conservé, ou cassette AAA (Beyer, 1997), qui est composé de plusieurs motifs protéiques supposés être responsables de l'activité ATPase chez tous les membres de la famille des AAA.

Tableau 4  
Mutations dans SPG4 chez les patients atteints de PSF-AD

Famille	Localisation	Mutation <sup>a</sup>	Changement d'acide aminé <sup>b</sup>	Conséquence
624	exon 7	1 210 C → G	S362C	faux sens
4 014	exon 11	1 468 G → A	C448Y	faux sens
A	exon 11	1 520 delT	466STOPcodon	non sens
618	exon 13	1 620 C → T	R499C	faux sens
2 992	intron 15	1 813-2a → g	Δ aa564 → aa576 (PTC+7 aa)	perte de l'exon 16 + décalage
5 226	intron 15	1 813-2a → g	Δ aa564 → aa576 (PTC+7 aa)	perte de l'exon 16 + décalage
5 330	intron 15	1 813-2a → g	Δ aa564 → aa576 (PTC+7 aa)	perte de l'exon 16 + décalage

<sup>a</sup> Les positions en nt font référence à la séquence de l'ADNc SPG4.

<sup>b</sup> Les positions en aa font référence à la séquence de la spastin.

Les bases des exons sont indiquées en majuscules, celles des introns en minuscules.

PTC+7 aa = "premature termination codon" à 7 aa en aval de l'exon 16.

En addition à ces cinq mutations décrites précédemment, des recherches de mutations hétérozygotes réalisées sur des patients atteints de PSF-AD issus de 36 autres familles ont permis de mettre en évidence 34 autres mutations altérant ou susceptibles d'altérer le produit d'expression du gène SPG4.

- 5 Les caractéristiques de ces 34 autres mutations sont résumées dans le tableau 5 ci-après dans lequel ont été également insérées les cinq premières mutations précédemment citées.

**Tabl au 5**  
**Mutations dans SPG4 chez les patients atteints de PSF-AD**

Famille	Localisation	Mutation <sup>a</sup>	Changement d'acide aminé <sup>b</sup>	Conséquence
624	exon 7	1210 C → G	S362C	faux sens
6958	exon 8	1233 G → A	G370R	faux sens
214	exon 8	1267 T → G	F381C	faux sens
1002	exon 8	1283 T → G	N386K	faux sens
027	exon 8	1288 A → G	K388R	faux sens
019	exon 10	1401 C → G	L426V	faux sens
4014	exon 11	1468 G → A	C448Y	faux sens
148	exon 11	1504 G → T	R460L	faux sens
618	exon 13	1620 C → T	R499C	faux sens
636	exon 15	1788 G → A	D555N	faux sens
627	exon 15	1792 C → T	A556V	faux sens
2971	exon 3	702 C → T	Q193STOP	non sens
3655	exon 5	873 A → T	K229STOP	non sens
1010	exon 5	907 C → A	S261STOP	non sens
3938	exon 5	932 C → G	Y269STOP	non sens
6922	exon 10	1416 C → T	R431STOP	non sens
616	exon 10	1416 C → T	R431STOP	non sens
605	exon 15	1809 C → T	R562STOP	non sens
030	exon 2	578-579insA	PTC + 2 aa	décalage + non sens
615	exon 5	852del11	PTC + 18 aa	décalage + non sens
042	exon 5	882-883insA	PTC + 12 aa	décalage + non sens
032	exon 5	906delT	PTC + 17 aa	décalage + non sens
189	exon 9	1299delG	PTC + 3 aa	décalage + non sens
3686	exon 9	1340del5	PTC + 35 aa	décalage + non sens
625	exon 9	1340del5	PTC + 35 aa	décalage + non sens
A	exon 11	1520delT	PTC + 7 aa	décalage + non sens
115	exon 12	1574delGG	PTC + 2 aa	décalage + non sens
3266	exon 13	1634del22	PTC + 18 aa	décalage + non sens
149	exon 14	1684-1685insTT	PTC + 9 aa	décalage + non sens
645	exon 14	1685del4	PTC + 7 aa	décalage + non sens
029	intron 4	808-2 a → g	?	mutation site d'épissage
162	intron 6	1129+2 t → g	?	mutation site d'épissage
125	intron 7	1223+1 g → t	?	mutation site d'épissage
143	intron 8	1299+1 g → a	?	mutation site d'épissage
1620	intron 11	1538+5 g → a	(PTC + 6 aa)	perte de l'exon 11 + décalage
1006	intron 11	1538+3 del4	?	mutation site d'épissage
1605	intron 13	1661+1 g → t	?	mutation site d'épissage
1012	intron 13	1662-2 a → t	?	mutation site d'épissage
1626	intron 15	1812+1 g → a	Δ aa564 → aa576 (PTC+7 aa)	mutation site d'épissage
2992	intron 15	1813-2 a → g	Δ aa564 → aa576 (PTC+7 aa)	perte de l'exon 16 + décalage
5226	intron 15	1813-2 a → g	Δ aa564 → aa576 (PTC+7 aa)	perte de l'exon 16 + décalage
5330	intron 15	1813-2 a → g	Δ aa564 → aa576 (PTC+7 aa)	perte de l'exon 16 + décalage
1611	intron 16	1813-2 a → g	?	mutation site d'épissage
		1853+1 g → a		

<sup>a</sup> Les positions en nt font référence à la séquence de l'ADNc SPG4. <sup>b</sup> Les positions en aa font référence à la séquence de la spastin. Les bases des exons sont indiquées en majuscules, celles des introns en minuscules. PTC+n aa - "premature termination codon" à n acide aminé en aval de la mutation.

### Exemple 5 : Analyse de la séquence protéique de la spastin

La phase ouverte de lecture de SPG4 code pour une protéine de 616 aa que nous avons nommée spastin et dont le poids moléculaire est d'environ 67,2 kDaltons (kD). La comparaison de cette séquence en acides aminés dans les bases de données protéiques à l'aide des programmes BLAST a permis de mettre en évidence une zone de forte homologie avec plusieurs membres de la famille des AAA à l'extrémité C-terminale de la spastin. Les motifs « types » de la famille des AAA, englobés dans la cassette AAA, sont localisés entre les positions aa342 et aa599 (voir Fig. 2) d'après les comparaisons de séquence dans les bases de données de domaines protéiques ProDom et Prosite. Les trois domaines types conservés, dont les motifs A et B de Walker ainsi que le motif consensus minimal des protéines AAA sont respectivement situés au sein de la cassette AAA aux positions aa382-389, aa437-442 et aa480-498 (Fig. 2). Le motif A de Walker, « GPPGNGKT » appelé également p-loop (ou boucle-p) qui correspond au domaine de fixation de l'ATP et le motif B « IIFIDE » sont très conservés entre tous les membres de la famille des AAA incluant la spastin.

La comparaison des cassettes AAA présentes dans 150 protéines de cette famille d'ATPase issues d'organismes très éloignés dans l'évolution a permis de classifier cet ensemble de protéines en plusieurs sous-groupes, en fonction du nombre de cassettes AAA identifiés (1 ou 2) et des homologies de séquence entre ces différentes cassettes (Beyer, 1997). Parmi toutes les protéines de la famille des AAA, la spastin présente une plus forte homologie avec une sous-classe particulière des AAA, et plus spécifiquement avec les protéines suivantes dont la plupart ont été identifiées grâce au séquençage complet du génome de l'organisme considéré : deux protéines de *Caenorhabditis elegans* O16299 et Q18128, deux sous-unités du protéasome 26S de *Saccharomyces cerevisiae* Yta6p (Q02845) et TBP6 (P40328) (Schnall et al., 1994), une sous-unité du protéasome de *Schizosaccharomyces pombe* (O43078), les protéines SAP1 (P39955) et END13 (P52917) de *S. cerevisiae* et la protéine murine SKD1 (P46467) (Perier et al., 1994). L'alignement multiple de ces 8 protéines avec la spastin est représenté sur la Fig. 4A. Sur les 257 acides aminés englobant la cassette AAA (positions aa342-599), la spastin présente une identité de séquence de 52 %, 51 % et 50 % avec la protéine de levure Yta6p (Q02845), la protéine de nématode O16299 et la protéine de levure TBP6 (P40328) respectivement. Des résultats similaires ont été obtenus par l'analyse de la séquence protéique de la spastin dans la base de données ProDom qui a montré l'existence de trois domaines d'homologie (nommés 92, 179 et 6226 et correspondant aux positions aa342-409, aa411-509 et aa512-599) trouvés dans les sous-unités putatives du protéasome 26S

de levure. En outre, les membres de ce sous-groupe des AAA contiennent le plus souvent des motifs de type leucine-zipper dont deux ont pu être détectés dans la séquence protéique de la spastin aux positions aa50-78 et aa508-529 par l'analyse de la séquence dans la base de données Prosite (voir Fig. 2). Cette analyse a également pu prédire la

5 présence d'un motif de dimérisation de type hélice-boucle-hélice (« helix-loop-helix ») situé entre les positions aa478 et aa486.

La comparaison de la séquence protéique de la spastin avec celles d s métalloprotéases mitochondriales comme les protéines de levure AFG3, RCA1 et YME1, ainsi que la paraplegin qui est impliquée dans une forme rare de PSF-AR montre que

10 l'homologie entre ces cinq membres de la famille des AAA est limitée à la région de 257aa englobant la cassette AAA (Fig. 4B). Dans cette région, l'identité de séquence entre la spastin et la paraplegin n'est que de 29 % alors que la paraplegin et la protéine de levure AFG3 sont identiques à 57 % sur cette même portion de la séquence protéique. Cette comparaison de séquence suggère que la spastin n'appartient pas au même sous-

15 groupe des AAA que la paraplegin et autres métalloprotéases mitochondriales. De plus, l'analyse informatique de la séquence de la spastin avec le programme PSORT II qui permet de prédire la localisation sub-cellulaire des protéines semble indiquer que le spastin est une protéine nucléaire. Un éventuel signal de localisation nucléaire (NLS pour « Nuclear Localization Signal »), RGKKK, a été mis en évidence entre les positions aa7 et

20 aa11 alors qu'aucun peptide signal caractéristique d'un import dans la mitochondrie n'a pu être décelé, contrairement à ce qui avait été observé pour la paraplegin.

#### Exemple 6 : Profils d'expression de SPG4 et de son orthologue murin Spg4

La comparaison de la séquence nucléique de SPG4 dans les banques de données d'EST a permis de détecter plusieurs ESTs humains, murins et de rat présentant une forte

25 homologie avec SPG4. Les clones d'ADNc de blastocyste et d'embryon E8 de souris correspondant à deux des ESTs murins, AA560327 et AA107866, ont été obtenus du consortium IMAGE et séquencés intégralement. L'assemblage des séquences de ces clones d'ADNc a permis de reconstituer une séquence consensus de 1689 pb incluant une phase ouverte de lecture incomplète de 1514 pb. La comparaison entre l'ADNc SPG4

30 humain et cet ADNc de souris a montré qu'il manque au transcrit murin environ 460 pb à l'extrémité 5' dont le codon d'initiation de la traduction. La phase de lecture ouverte de souris est suivie d'une région 3' non codante (3' UTR) de 175 pb contenant un site de polyadénylation situé à ~20 pb en amont de la queue polyA (Fig. 5). La séquence nucléique de SPG4 et la séquence protéique de la spastin humaine présentent

35 respectivement une identité de 89 % (entre les positions nt 460 et nt 1982) et de 96 %

(entre les positions aa113 et aa616) avec les séquences de l'ADNc et de la protéine déduite de la souris. Ce degré important d'homologie permet d'affirmer que ce transcrit de souris correspond à l'orthologue murin de SPG4, qui a été donc baptisé Spg4.

L'hybridation de northern blots comprenant les ARNm de divers tissus humains et murins (Clontech) avec les clones d'ADNc SPG4 et Spg4 n'a pas donné de résultats probants excepté une très faible bande correspondant à un transcrit de 2,5 kb dans le testicule de souris après 10 jours d'exposition. En raison du faible niveau d'expression de ce gène, les profils d'expression de SPG4 et Spg4 ont été déterminés par des expériences de PCR sur des collections d'ADNc normalisées provenant de divers tissus adultes et fœtaux (voir Fig. 6A à 6C). Le gène murin Spg4 est exprimé de façon ubiquitaire dans les tissus adultes de souris ainsi que du stade E7 au stade E17 de l'embryon de souris (Fig. 6A). Une plus forte expression de Spg4 a été détectée dans le foie, le muscle squelettique et les testicules, ainsi qu'au stade E15 de l'embryon. L'expression précoce de Spg4 au cours du développement embryonnaire a été confirmée par la présence d'ESTs provenant de banques d'ADNc de blastocyste, d'embryon E8 et de carcinome embryonnaire dans les banques de données publiques d'ESTs. Le gène humain SPG4 est lui aussi exprimé de façon ubiquitaire dans les tissus adultes (Fig. 6B) et fœtaux (Fig. 6C), avec une expression peut-être plus marquée dans le cerveau fœtal.

Exemple 7 : Pas de défaut de la phosphorylation oxydative dans la PSF-AD liée au locus SPG4

Afin de déterminer si des mutations de la spastin induisaient un défaut de la phosphorylation oxydative (OXPHOS) dans la mitochondrie, à l'image de ce qui avait été observé pour la paraplegin, une biopsie musculaire a été réalisée sur un patient d'une des familles de PSF-AD liée au locus SPG4. Les analyses morphologiques et histoenzymatiques de cette biopsie de muscle n'ont pas révélé de fibre musculaire de type RRF (pour « ragged red fibers »), caractéristique des défauts OXPHOS dans la mitochondrie. Le fait que toutes les fibres musculaires apparaissent normales ainsi que la prédiction d'une localisation nucléaire de la spastin semblent indiquer que la PSF-AD liée au locus SPG4 n'est pas une maladie mitochondriale de type OXPHOS, par opposition à la PSF-AR liée au locus SPG7.

Par une approche de clonage positionnel basée sur le séquençage d'une région de 1,5 Mb, nous avons identifié le gène SPG4 (ou SPAST) responsable de la forme la plus fréquente de PSF-AD, préalablement localisé sur les bandes chromosomiques 2p21-p22. Trente-neuf mutations altérant ou susceptibles d'altérer le produit du gène, nommé

spastin, ont pu être détectées chez les individus atteints de quarante-et-une familles de PSF-AD présentant une liaison au locus SPG4. La spastin est un nouveau membre de la famille des protéines AAA, dont la localisation semble être nucléaire et qui présente une forte homologie avec les sous-unités du protéasome 26S de levure. En dépit d'une

5 grande homologie restreinte à un domaine de 230 à 250 aa, dite cassette AAA, les nombreux membres de cette famille protéique peuvent participer à des mécanismes cellulaires très variés comme le transport de protéines au sein de vésicules, la régulation du cycle cellulaire, la biogénèse des organelles, le contrôle de la transcription, ....

10 Toutefois, tous ces mécanismes cellulaires impliquent l'assemblage, la fonction ou la dégradation de complexes protéiques, ce qui suggère que les membres de la famille des AAA sont des protéines dites « chaperons ».

## Références

- Barany, F., (1991), Proc. Natl. Acad. Sci. USA, 88, 189-193.
- Beyer, A. Sequence analysis of the AAA protein family. Protein Sci. 6, 2043-2058 (1997).
- Bodansky M., Principles of peptide synthesis, (1984).
- 5 Bruyn, R.P.M. & Scheltens, P.H. Hereditary spastic paraparesis (Strumpell-Lorrain) in Handbook of clinical neurology Vol. 15 (ed. de Jong, J.M.B.V.) 301-318 (Elsevier Science Publishers B.V., 1991).
- Buckholz, R.G. Curr. Op. Biotechnology 4 : 538-542, 1993.
- Burg, J.L. et al. (1996), Mol. and Cell. Probes, 10, 257-271.
- 10 Carter, B.J. Curr. Op. Biotechnology 3 : 533-539, 1993.
- Casari, G. et al. Spastic paraplegia and OXPHOS impairment caused by mutations in Paraplegin, a nuclear-encoded mitochondrial metalloprotease. Cell 93, 973-983 (1998).
- Cherif D., Julier, C., Delattre, O., Derré, J., Lathrop, G.M., and Berger, R. Proc. Natl. Acad. Sci. USA. 87 : 6639-6643, 1990.
- 15 Chu, B.C.F. et al. (1986), Nucleic Acids Res., 14, 5591-5603.
- Chumakov, I., Rigault, P., Guillou, S., Ougen, P., Billault, A., Guasconi, G., Gervy, P., Le Chumakov, I.M., Rigault, P., Le Gall, I., et al. Nature 377 : 175-183, 1995.
- Confalonieri, F. & Duguet, M. A 200-amino acid ATPase module in search of a basic function. BioEssays 17, 639-650 (1995).
- 20 Duck, P. et al. (1990), Biotechniques, 9, 142-147.
- Durr, A. et al. Phenotype of autosomal spastic paraplegia linked to chromosome 2. Brain 119, 1487-1496 (1996).
- Edwards, C.P., and Aruffo, A. Curr. Op. Biotechnology 4 : 558-563, 1993.
- Epstein, A. Médecine/Sciences 8 : 902-911, 1992.
- 25 Erlich, H.A., (1989), New York : Stockton Press.
- Guatelli J.C. et al. Proc. Natl. Acad. Sci. USA 87: 1874-1878, 1990 et al. Cell 85 : 281-290, 1996.
- Fink, J.K. et al. Autosomal dominant familial spastic paraplegia : tight linkage to chromosome 15q. Am. J. Hum. Genet. 56, 188-192 (1995).
- 30 Hazan, J., Lamy, C., Melki, J., Munnich, A., de Recondo, J., & Weissenbach, J. Autosomal dominant familial spastic paraplegia is genetically heterogeneous and one locus maps to chromosome 14q. Nature Genet. 5, 163-167 (1993).
- Hazan, J. et al. Linkage of a new locus for autosomal dominant familial spastic paraplegia to chromosome 2p. Hum. Mol. Genet. 3, 1569-1573 (1994).

- Hedera, P. et al. Novel locus for autosomal dominant hereditary spastic paraplegia, on chromosome 8q. *Am. J. Hum. Genet.* 64, 563-569 (1999).
- Heinzlef, O. et al. Mapping of a complicated familial spastic paraplegia to locus SPG4 on chromosome 2p. *J. Med. Genet.* 35, 89-93 (1998).
- 5 Hentati, A. et al. Linkage of a locus for autosomal dominant familial spastic paraplegia to chromosome 2p markers. *Hum. Mol. Genet.* 3, 1867-1871 (1994).
- The Hereditary Spastic Paraplegia Working Group. Hereditary spastic paraplegia : advances in genetic research. *Neurology* 46, 1507-1514 (1996).
- Innis, M.A. et al. (1990), Academic Press.
- 10 Jouet, M. et al. X-linked spastic paraplegia (SPG1), MASA syndrome and X-linked hydrocephalus result from mutations in the L1 gene. *Nature Genet.* 7, 402-407 (1994).
- Kievitis, T. et al. (1991), *J. Virol. Methods*, 35, 273-286.
- Köhler et Milstein. *Nature* 256, 495-497, 1975.
- Kwoh, D.Y. et al. (1989), *Proc. Natl. Acad. Sci. USA*, 86, 1173-1177.
- 15 Landegren U., Kaiser R., Sanders J. & Hood L. *Science* 241 : 1077-1080, 1988.
- Lizardi, P.M. et al. (1988), *Bio/technology*, 6, 1197-1202.
- Luckow, V.A. (1993), *Curr. Op. Biotechnology* 4, 564-572.
- Matthews, J.A. et al. (1988), *Anal. Biochem.*, 169 : 1-25.
- Miele, E.A. et al. (1983), *J. Mol. Biol.*, 171 : 281-295.
- 20 Nielsen, J.E. et al. CAG repeat expansion in autosomal dominant pure spastic paraplegia linked to chromosome 2p21-24. *Hum. Mol. Genet.* 6, 1811-1816 (1997).
- Olins, P.O., and Lee, S.C. *Curr. Op. Biotechnology* 4 : 520-525, 1993.
- Osoegawa, K. et al. An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 52, 1-8 (1998).
- 25 Perier, F. et al. Identification of a novel mammalian member of the NSF/CDC48p/Pas1p/TBP-1 family through heterologous expression in yeast. *FEBS lett.* 351, 286-290 (1994).
- Perricaudet, M., Stratford-Perricaudet, L. and Briand, P. *La Recherche* 23 : 471-473, 1992.
- 30 Polo, J.M., Calleja, J., Combarros, O. & Berciano, J. Hereditary ataxias and paraplegias in Cantabria, Spain. An epidemiological and clinical study. *Brain* 114, 855-866 (1991).
- Reid, E. Pure hereditary spastic paraplegia. *J. Med. Genet.* 34, 499-503 (1997).
- Rohlmann, A., Gotthardt, M., Willnow, T.E., Hammer, R.E., and Herz, J. *Nature Biotech.* 14 : 1562-1565, 1996.
- 35 R lfs, A. et al. (1991), Berlin : Springer-Verlag.

- Sambrook, J., Fritsch, E.F., and Maniatis, T. Molecular cloning : a laboratory manual.
- Saugier-Weber, P. et al. X-linked spastic paraplegia and Pelizaeus-Merzbacher disease are allelic disorders at the proteolipid protein locus. *Nature Genet.* 6, 257-262 (1994).
- Schnall, R. et al. Identification of a set of yeast genes coding for a novel family of putative
- 5 ATPases with high similarity to constituents of the 26S protease complex. *Yeast* 10, 1141-1155 (1994).
- Scott, W.K. et al. Locus heterogeneity, anticipation, and reduction of the chromosome 2p minimal candidate region in autosomal dominant familial spastic paraplegia. *Neurogenetics* 1, 95-102 (1997).
- 10 Sec. Ed. Cold Spring Harbor Lab., Cold Spring Harbor, New York.
- Segev, D., (1992), Kessler C. Springer Verlag, Berlin, New-York, 197-205.
- Skre, H. Hereditary spastic paraplegia in Western Norway. *Clin. Genet.* 6, 165-183 (1974).
- Stone, B.B. et al. (1996). *Mol. and Cell. Probes*, 10 : 359-370.
- Stewart J.M. et Yound J.D., solid phase peptides synthesis, Pierce Chem. Company,
- 15 Rockford, 111, 2ème éd., (1984).
- Suggs S.V., Wallace R.B., Hirose T., Kawashima E.H. and Itakura K. *PNAS* 78 : 6613-6617, 1981.
- Temin, H.M. Retrovirus vectors for gene transfer. In Kucherlapati R., ed. *Gene Transfer*, New York, Plenum Press, 149-187, 1986.
- 20 Walker G.T., Fraiser M.S., Schram J.L., Little M.C., Nadeau J.G., & Malinowski D.P. *Nucleic Acids Res.* 20 : 1691-1696, 1992.
- Werderlin, L. Hereditary ataxias. Occurrence and clinical features. *Acta Neurol. Scand.* 73 (Suppl. 106) (1986).
- Woo S.L.C. *Methods Enzymol.* 68 : 389, 1979.

## REVENDICATIONS

1. Acide nucléique purifié ou isolé du gène SPG4, caractérisé en ce qu'il comprend une séquence choisie parmi le groupe comprenant :
  - 5 a) la séquence SEQ ID No. 1, la séquence SEQ ID No. 2, la séquence SEQ ID No. 72, la séquence la séquence SEQ ID No. 106 ou la séquence d'au moins 15 nucléotides consécutifs d'une de ces séquences ;  
les séquences nucléiques homologues ou variantes des séquences SEQ ID No. 1, SEQ ID No. 2, SEQ ID No. 72 ou SEQ ID No. 106 ; et
  - c) la séquence complémentaire ou la séquence de l'ARN correspondant aux séquences telles que définies en a) et b).
2. Acide nucléique purifié ou isolé selon la revendication 1, à l'exception de l'acide nucléique identifié dans la banque de données GenBank sous le numéro d'accension AB029006.
- 15 3. Acide nucléique purifié ou isolé selon la revendication 1 ou 2, caractérisé en ce qu'il comprend au moins une séquence d'au moins 15 nucléotides consécutifs, du fragment nt 714-809, extrémités incluses, de la séquence SEQ ID No. 2, de sa séquence complémentaire ou de la séquence de son ARN correspondant.
4. Acide nucléique purifié ou isolé selon l'une des revendications 1 à 3,  
20 caractérisé en ce qu'il comprend une mutation correspondant à un polymorphisme naturel chez l'Homme.
5. Sonde ou amorce, caractérisée en ce qu'elle comprend une séquence d'un acide nucléique selon l'une des revendications 1 à 4.
6. Sonde ou amorce selon la revendication 5, caractérisée en ce qu sa  
25 séquence est choisie parmi les séquences SEQ ID No. 4 à SEQ ID No. 71.
7. Site accepteur ou donneur d'épissage, caractérisé en ce qu'il comprend une séquence d'un acide nucléique selon la revendication 1 choisie parmi les séquences SEQ ID No. 74 à SEQ ID No. 105.
8. Méthode pour le criblage de banques d'ADNc ou d'ADN génomique ou  
30 pour le clonage d'ADNc ou génomique isolé codant pour la spastin, caractérisée en ce qu'elle met en œuvre une séquence nucléique selon l'une des revendications 1 à 7.
9. Méthode selon la revendication 8, pour l'identification de la séquence génomique ou de l'ADNc du gène SPG4 de mammifère, notamment de souris.

10. Méthode pour l'identification de mutation portée par le gène SPG4 humain, caractérisée en ce qu'elle met en œuvre une séquence nucléique selon l'une des revendications 1 à 7.

5 11. Méthode selon la revendication 10, pour l'identification de mutation responsable de la paraplégie spastique familiale autosomique dominante.

12. Méthode pour l'identification des séquences d'acide nucléique promotrices et/ou régulatrices de l'expression du gène SPG4, caractérisée en ce qu'elle met en œuvre une séquence nucléique selon l'une des revendications 1 à 7.

10 13. Acide nucléique identifié par une méthode selon l'une des revendications 9 à 12.

14. Polypeptide codé par un acide nucléique selon l'une des revendications 1 à 4 et 13.

15 15. Polypeptide selon la revendication 14, de préférence à l'exception du peptide de 584 acides aminés dont la séquence est identifiée dans la banque de données GenBank sous le numéro d'accèsion AB029006.

16. Polypeptide selon la revendication 14 ou 15, caractérisé en ce qu'il comprend une séquence d'acides aminés choisie parmi le groupe comprenant :

20 a) la séquence SEQ ID No. 3, la séquence SEQ ID No. 73, la séquence SEQ ID No. 107 ou la séquence d'au moins 10 acides aminés consécutifs d'une de ces séquences ; et  
b) les séquences homologues ou variantes des séquences SEQ ID No. 3, SEQ ID No. 73 ou SEQ ID No. 107.

17. Polypeptide selon la revendication 14 ou 15, caractérisé en ce qu'il comprend la séquence d'au moins 8 acides aminés consécutifs de la séquence du fragment aa 197-228, extrémités incluses, de la séquence SEQ ID No. 3.

25 18. Polypeptide selon la revendication 14 ou 15, caractérisé en ce qu'il comprend une séquence d'acides aminés choisie parmi le groupe comprenant la séquence SEQ ID No. 3, la séquence SEQ ID No. 73, la séquence SEQ ID No. 107, lesquelles séquences portant au moins une des mutations correspondant à un polymorphisme naturel chez l'Homme, et les séquences de leurs fragments d'au moins 10  
30 acides aminés consécutifs.

19. Vecteur de clonage et/ou d'expression contenant une séquence d'acide nucléique selon l'une des revendications 1 à 4, et 13.

20. Vecteur selon la revendication 19, caractérisé en ce qu'il comporte les éléments nécessaires à son expression dans une cellule hôte.

35 21. Cellule hôte transformée par un vecteur selon la revendication 19 ou 20.

22. Mammifère, excepté l'Homme, caractérisé en ce qu'il comprend une cellule selon la revendication 21.

23. Mammifère, excepté l'Homme, selon la revendication 22, comprenant une cellule transformée, caractérisé en ce que la séquence de l'un au moins des deux allèles du gène SPG4 contient une au moins des mutations correspondant à un polymorphisme naturel chez l'Homme ou identifiées par une méthode selon la revendication 10 ou 11.

24. Utilisation d'une séquence d'acide nucléique selon l'une des revendications 5, 6 et 13 comme sonde ou amorce, pour la détection et/ou l'amplification de séquences d'acide nucléique.

25. Utilisation d'une séquence d'acide nucléique selon l'une des revendications 1 à 7, et 13, pour le criblage de banque génomique ou d'ADNc.

26. Utilisation d'une séquence d'acide nucléique selon l'une des revendications 1 à 4, et 13, pour la production d'un polypeptide recombinant ou synthétique.

27. Méthode de production d'un polypeptide recombinant, caractérisée en ce que l'on cultive une cellule transformée selon la revendication 21 dans des conditions permettant l'expression dudit polypeptide recombinant et que l'on récupère ledit polypeptide recombinant.

28. Polypeptide, caractérisé en ce qu'il est obtenu par une méthode selon la revendication 27.

29. Anticorps mono- ou polyclonaux ou leurs fragments, anticorps chimériques ou immunoconjugués, caractérisés en ce qu'ils sont capables de reconnaître spécifiquement un polypeptide selon l'une des revendications 14 à 18, et 28.

30. Méthode pour la détection et/ou la purification d'un polypeptide selon l'une des revendications 14 à 18, et 28, caractérisée en ce qu'elle met en oeuvre un anticorps selon la revendication 29.

31. Méthode de diagnostic génotypique de la PSF-AD associée au gène SPG4, caractérisée en ce que l'on met en oeuvre une séquence d'acide nucléique selon l'une des revendications 1 à 7, et 13.

32. Méthode de diagnostic génotypique de PSF-AD associée à la présence d'au moins une mutation sur une séquence du gène SPG4 à partir d'un prélèvement biologique d'un patient, caractérisée en ce qu'elle comporte les étapes suivantes :

a) le cas échéant, isolement de l'ADN génomique à partir de l'échantillon biologique à analyser, ou obtention d'ADNc à partir de l'ARN de l'échantillon biologique ;

b) amplification spécifique de ladite séquence d'ADN du gène SPG4 susceptible de contenir une mutation à l'aide d'amorces selon l'une des revendications 5 et 6 ou d'un acide nucléique selon la revendication 13 ;

5 c) analyse des produits d'amplification obtenus et comparaison de leur séquence avec la séquence normale correspondante du gène SPG4.

33. Méthode de diagnostic de la PSF-AD associée à une expression anormale d'un polypeptide codé par le gène SPG4, caractérisée en ce que l'on met en contact un ou des anticorps selon la revendication 29 avec le matériel biologique à tester, dans des conditions permettant la formation éventuelle de complexes immunologiques spécifiques dudit polypeptide et le ou lesdits anticorps, et en ce que l'on détecte et/ou quantifie les complexes immunologiques éventuellement formés.

34. Méthode de sélection d'un composé chimique ou biochimique capable d'interagir directement ou indirectement avec un polypeptide selon l'une des revendications 14 à 18, et 28, ou avec un acide nucléique selon l'une des revendications 15 1 à 7, et 13, et/ou permettant de moduler l'expression ou l'activité de ces polypeptides, caractérisée en ce qu'elle comprend la mise en contact d'une séquence d'acide nucléique selon l'une des revendications 1 à 7, et 13, d'un polypeptide selon l'une des revendications 14 à 18, et 28, d'un vecteur selon l'une des revendications 19 et 20, d'une cellule selon la revendication 21, d'un mammifère selon l'une des revendications 22 et 23 20 ou d'un anticorps selon la revendication 29, avec un composé candidat et, la détection d'une modification de l'activité dudit polypeptide.

35. Utilisation d'une séquence d'acide nucléique selon l'une des revendications 1 à 7, et 13, d'un polypeptide selon l'une des revendications 14 à 18, et 28, d'un vecteur selon l'une des revendications 19 et 20, d'une cellule selon la revendication 21, d'un 25 mammifère selon l'une des revendications 22 et 23 ou d'un anticorps selon la revendication 29, pour l'étude de l'expression ou de l'activité du gène SPG4.

36. Trousse ou nécessaire pour le diagnostic, caractérisée en ce qu'elle comprend au moins un composé choisi parmi le groupe de composés suivant :

- 30 a) un acide nucléique selon l'une des revendications 5 et 6 ; et  
b) un anticorps selon la revendication 29.

1/7

FIGURE 1A

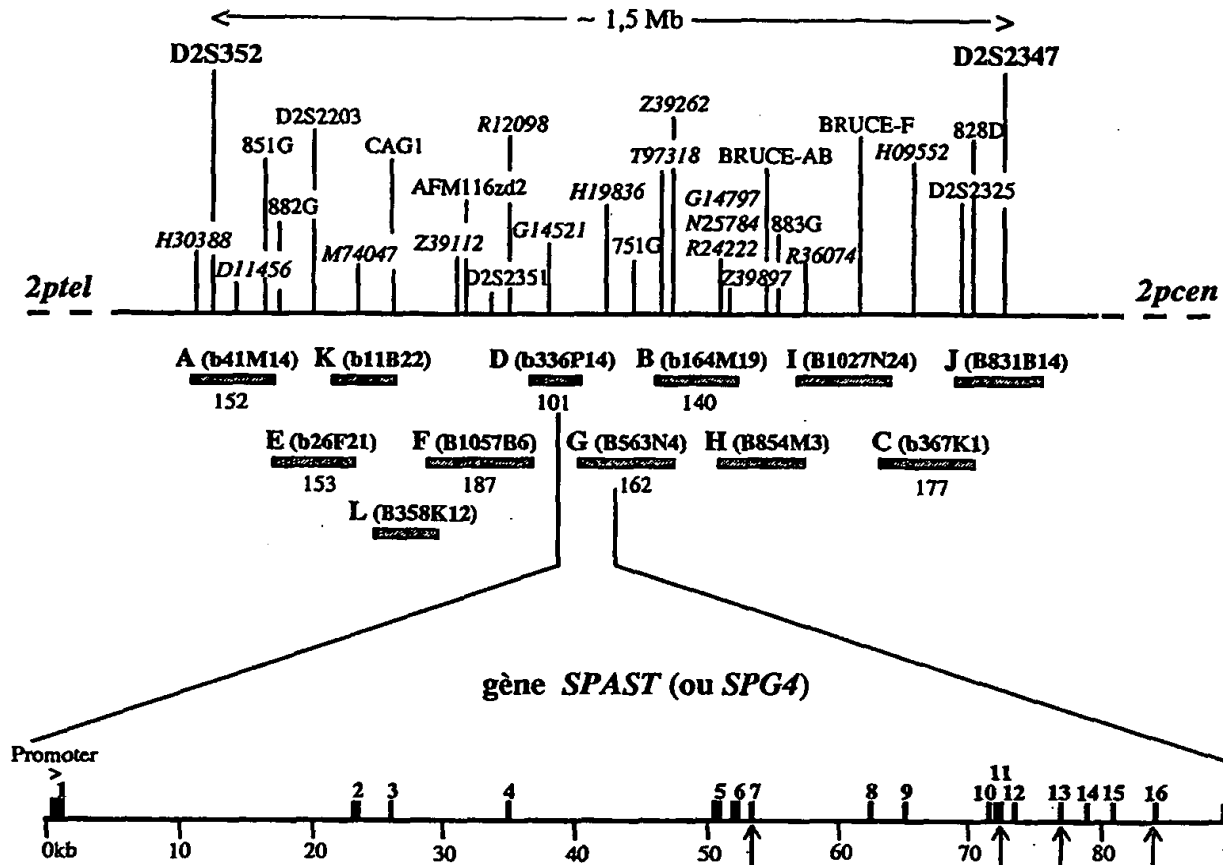
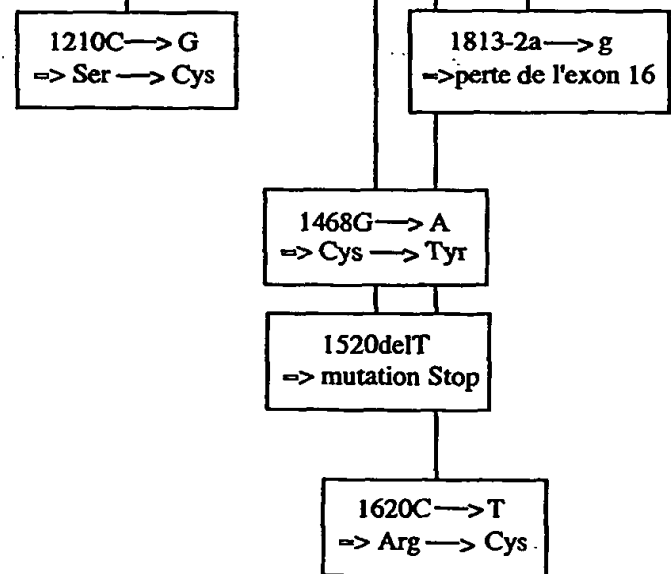


FIGURE 1B

FIGURE 1C





.

.

.

.

2/7

1 GCTCCTGAGACCGGCGGGCACACGGGGGTCTGTGGCCCCCGCCGTAGCAGTGGCTGCCGCCGTTCGGTTCCCGCTCGGTCTGCGGGAGGCGGG 95  
 1 TTATGGCGGCGGCGGCAGTGAAGAGCTGTGAATGAATTCTCCGGGTGGACGAGGGAAGAAGAAAGCTCCGGCGGCGCCAGCAACCCGGTGCCTCC 190  
 1 M N S P G G R G K K K G S G G A S N P V P P  
 CAGGCTCCGCCCCCTTGCTGGCCCCCGCCCTCCCGCCCGGGCGGCCCCCTCCGCCGAGTCCGCCGATAAGCGGAACCTGTACTATTCT 285  
 23 R P P P P C L A P A P P A A G P A P P P E S P H K R N L Y Y F S  
 CCTACCCGCTGTTTGTAGGCTTCGCGCTGCTGCGTTTGGTTCGCTTCCACTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGCTTCTCCCGCGCC 380  
 55 Y P L F V G F A L L R L V A F H L G L L F V W L C Q R F S R A  
 CTCTGGCAGCCAAGAGGAGCTCCGGGGCGCGCCAGCTGCTCGGGCTCCGGCCCCGGCGCGGTGCCGGGCGGAGGCGGAGCGCGCTCCG 475  
 86 L M A A K R S S G A A P A P A S A S A P A P V P G G E A E R V R  
 AGTCTTCCACAAACAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGTGGACAGAAGGAGCAAGCTGTGGAATGGT 570  
 118 V F H K Q A F E Y I S I A L R I D E D E K A 2 G Q K E Q A V E W Y  
 ATAAGAAAGGTATTGAAGAACTGGAAAAAGGAATAGCTGTTATAGTTACAGGACAAGTGAACAGTGTGAAAGAGCTAGACGCCCTTCAAGCTAAA 665  
 150 K K G I E E L E K G I A V I V T G Q G 3 E Q C E R A R R L Q A K  
 ATGATGACTAATTTGGTTATGGCCAGGACCGCTTACAACCTTCTAGAGAAGATGCAACCACTTTTGCCATTTCCTAAGTCAAAACGGACGCTTA 760  
 181 M M T N L V M A K D R L Q L L E 4 K M Q P V L P F S K S Q T D V Y  
 TAATGACTACTAATTTGGCATGCGCAATGGACATCTCCAGTCAAGAGTGGAGCTGTCCAAAAAGAAAAGCCCTTAACACACACTAGTA 855  
 213 N D S T N L A C R N G H L Q S E 5 S G A V P K R K D P L T H T S N  
 ATTCAGTGCCTCGTTCAAAAACAGTTATGAAAACCTGGATCTGCAGGCTTTCAGGCCACCATAGAGCACCTAGTTACAGTGGTTTATCCATGGTT 950  
 245 S L P R S K T V M K T G S A G L S G H H R A P S Y S G L S M V  
 TCTGGAGTGAACAGGGATCTGGTCTGCTCTACCACTCTAAGTGTACTCCGAAAACAAATAGGACAAAATAAACCTTCTACCCCTACAACCTGC 1045  
 276 S G V K Q G S G P A T T H K G 6 T P K T N R T N K P S T P T A  
 TACTCGTAAGAAAAAGACTTGAAGAATTTTAGGAATGTGGACAGCAACCTTGTAACCTTATAATGAATGAAATGTGGACATGGAACAGCTG 1140  
 308 T R K K K D L K N F R N V D S N L A N L I M N E I V D N G 7 T A V  
 TTAATTTGATGATATAGCTGGTCAAGACTTGGCAAAACAGCATTGCAAGAAATGTTATTCTTCTCTGAGGCCTGAGTTGTTACAGGG 1235  
 340 K F D D I A G Q D L A K Q A L Q E I V I L P S L R P E L 8 F T G  
 CTTAGAGCTCCTGCCAGAGGGCTGTTACTCTTTGGTCCACCTGGGAATGGGAAGACAATGCTGCTAAAGCAGTAGCTGCAGAATCGAATGCAAC 1330  
 371 L R A P A R G L L L F G P P G N G K T M L A 9 K A V A A E S N A T  
 CTTCTTAATATAAGTGTGCAAGTTTAACTTCAAAATAGTGGGAGAGGAGAGAAATGGTGAGGGCTCTTTTGTGTGGCTCGAGAACTTC 1425  
 403 F N I S A S L T S K Y 10 G E G E K L V R A L F A V R E L Q  
 AACCTTCTATAATTTTATAGTGAAGTTGATAGCTTTTGTGTGAAAGAAGAGAAGGGGAGCAGCATGTAGTAGACGCCATAAACTGAATTT 1520  
 435 P S I I F I D 11 E V D S L L C E R R E G E H D A S R R L K T E F  
 CTAATAGAAATTTGATGGTGTACAGTCTGCTGGAGATGCACAGTACTTGTAAATGGGTGCAACTAATAGGCCACAAGAGCTGTATGAGGCTGTTCT 1615  
 466 L I E F D G V 12 Q S A G D R V L V M G A T T N R P Q E L D E A V L  
 CAGCGTTCATCAAAACGGGTATATGTCTTTACCAAAATGAGGAGTCAAGACTACTTTTGCTTAAAAATCTGTTATGTAACCAAGGAAGTCCAT 1710  
 498 R 13 R F I K R V Y V S L P N E E 14 R L L L L K N L L C K Q G S P L  
 TGACCCAAAAAGAACTAGCACAACTGCTAGTATGACTGATGGATACTCAGGAAGTGACCTAACAGCTTTGGCAAAAGATGAGCACTGGGTCTCT 1805  
 530 T Q K E L A Q L A R M 15 T D G Y S G S D L T A L A K D A A L G P  
 ATCCGAGTAACTAAAAACGAGGAGTGAAGATATGCTCGCCAGTGAGTGAAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAT 1900  
 561 I R E 16 L K P E Q V K N M S A S E M 17 R N I R L S D F T E S L K K I  
 AAAACGCAGCGTCAGCCCTCAAACCTTTAGAAGCGTACATACGTTTGAACAAAGGACTTTGGAGATACCAGTGTAAAGGAAATACCTTTGTAAACC 1995  
 593 K R S V S P Q T L E A Y I R W N K D F G D T T V \*  
 TGCAGAACATTTTACTTAAAAGAGGAAACACAAGATCTTCAATGAACGTCTCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTATAGAGTCT 2090  
 TACATATTTGTGCACCAAACTTGAAGATGAACAGAAAAACAGACTTAAACAAAAATATACAAATGCAAAATGTAATTTTGTGTTTAAAGCCTTGC 2185  
 CTTGATGGTCACAGTTATCCCAATGGACACTAAGTTAGAGCACAAACAACTGATTCTGGTCTTCTTTACCAATATAATCATAATGTAAATAAT 2280  
 AATTTGTATATTGTGTTGCAGATGAAAGTATCCAGGAACAGTGAATGGTAGAAGACACAAGAACATTTGTTGTTGTCTTCTGATGTTTTC 2375  
 TTAATAATAGTAATTTCTCTACTTTTCTTTCTACTGTTGCTTAACTACAGGTGATGGAAATGCCAAACACTCTTAAGTTTATTTCTTTTTC 2470  
 GTTTTATAAATTCAGTGTGCCAAATGAAACTTTTTTCTAAGTAACGTAAATAGGAAAAAGTTTATTTTGAGAGTTCTTCTTCAAAATCTACA 2565  
 GACATTAACAATTTGTTGTGTTCTTTTACCTTTTATTTTCTATTACCTTGCTACCAACAGTTTATAGATGCAATATAATAGCAAAAAAGCAAA 2660  
 TATGTTAAATAGAGAAGGTTTGAAGGTTTGAAGTTACTCTGTCATATAACATGTAGATCAGTCTTCATGTGACCTGCAGTATTTTTTTTCTAAT 2755  
 GTATTTGTGAGAAATCTGTTGTAGACTGTTAACTTCTTCTGATGGAATTTATTTCTGCAAGAATTATCTGATATTTAAGAGAGCCAAATTTTA 2850  
 ACTGCTGTGAAAAATGTTTCCAGTGCAAGAGAAGGGAAATACAGGAACATAAGACATTTCTAATTTATTGCTTATTACTTTCTAATTTTACAGGA 2945  
 TAATTATAAGCAAGTGAAGTACCATCTTTTATTCTTAATAATTATTAATCCCTTCAATGAAACTTTAAAAAACTGAATTTTATACATGGCAT 3040  
 ACAATTTTCTAGTTCCTTCTGCTTTTATTAACCTCAAAAGTTCTAGTTCTAGTCTGTTGATCTGCCCTTTGTTCTCCCAAAATGTACAGTAAT 3135  
 TCCATTTGTTGTATAAATATGCCTGGATTTTCATTATAAAAAATGTCATTGTAGGGAGTAGAGACTCATATCATGGCCTTTTAAATATTGTAATA 3230  
 AAGGCAATAGATATTTGCCCTTAGTTTACTGG 3263

FIGURE 2



FIGURE 3A

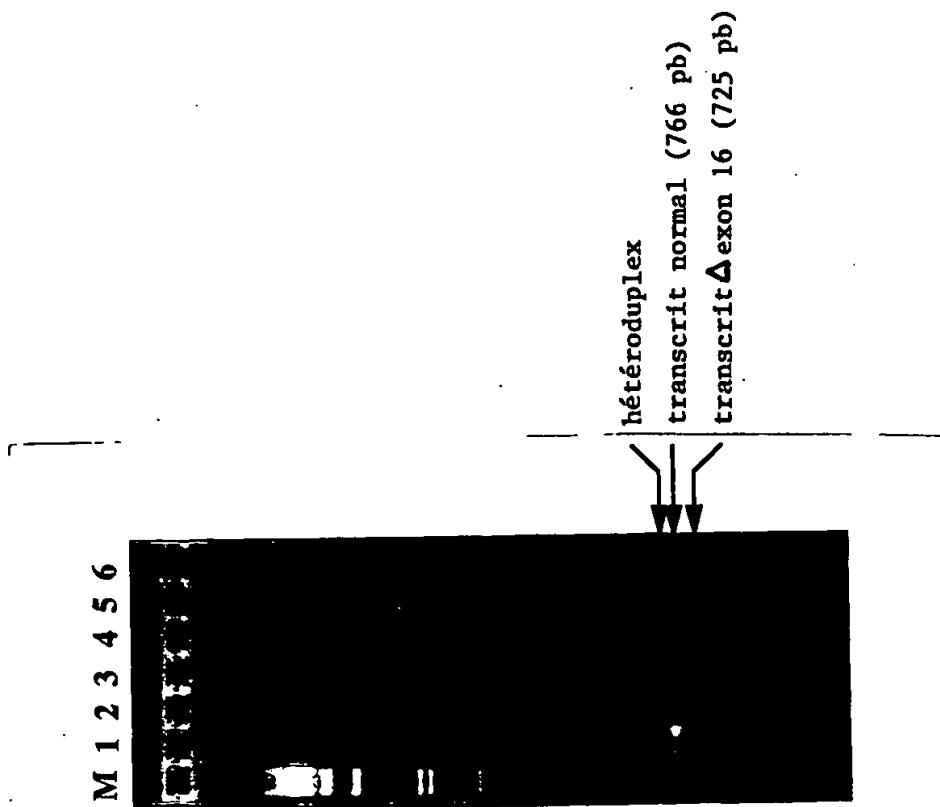
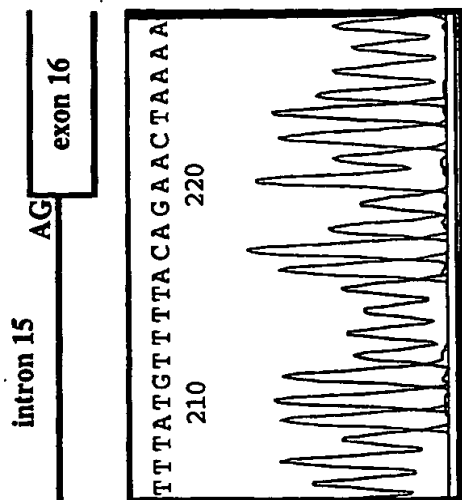
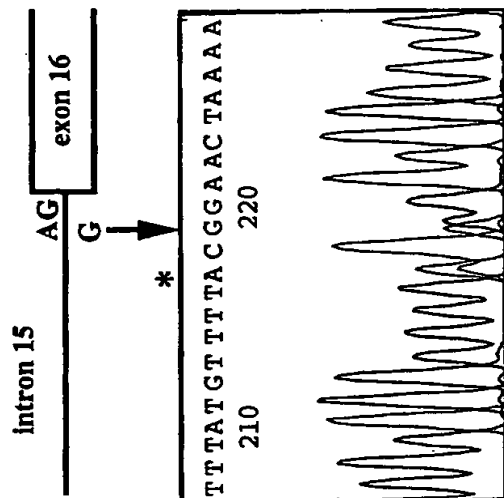


FIGURE 3B

Site accepteur d'épissage normal



Site accepteur d'épissage muté





4/7

[illegible]

**FIGURE 4A**



5/7

```

AFG3_YEAST ---MMHWORYARGAPE---SLTSL---FGKASRISTVKP-----VLRSEMPVHORLQTLSGLATR-----NTIHR--S
RCAL_YEAST -MLLLSWSRIATKVVRFPVFRSYGLTHIRSLHTQYRLNRLQENKSGNKNEDNEDAKLNKEIPTDEVEAIRKQVEKYIEQTKNRTIPANWK
PARAPLEGIN MAVLLLLLRRLRRGPGPGPH---PLMGPGFAMS-FCFPARFGRGEP-----YMASBPFGD--LAEAGGRALQ-----S
YME1_YEAST ---MNVSKILVSPVTTN-----VLRIFAPRLFOICASLHVOKKNA-----LRSKKFYRFYSEKNSCEMFP-----K
SPAST_HUMAN --MNSPGGRGKKKSGGCASN-----PVFPFPFPFCLAPAPPAAGPAPFP-----ESPFRKNLYYSYPLFVFGALLRLVAFHGLLFLVWLC

AFG3_YEAST TQIRSPHISWTEHLN-----ENRPNKE---GEGKNNGKNDNNSNKEGGRKNEFCSLG-----EYFESREFANTMFLTIGFTIIF
RCAL_YEAST EQRRKIDESIRLEDAVLKQESNRIQESRKEKEEENCPSKAKNSRTEQCYFECNNSRNIPPPPPPPPKPPLNDPSPNFVSKHVLFOIGLTFPL
PARAPLEGIN LQLELLTPTTEGINGLLKQH--LVQNEVRLWQLLGGTFYFHTSRLKQKNNKDKSKGKAPEEDE-----EERRRERDDQMYREKLRLTLV
YME1_YEAST KEADSSGRASNKST-----ISSIDMSQF--PFFSNTDKTKQANVAVSHAMLATHEQEA-----NKDLTSPDAQAAFYKLLQSNY
SPAST_HUMAN QRFSSRALMAAKRSSGAAPAP---ASASAPAPVF--GGEAERVVPHKQAPFYISIALRIDEDEKAG-----QKEQAVEWYKKGISELEKGIIV

AFG3_YEAST TLLTPSSMNSGDDSHRVLTFQDFKRYLEKCLVSKYIYVHKF-LVEAEVLN-----TKQVVSFTIGSV---DIFEEQ--MDQIQDLLNI
RCAL_YEAST LSFLLDLLNSLEEQS-EITWQDFREKLLAKGYVAKLIVNKS-MVKVMLNDNGKNQADNYGRNFYFTIGSI---DSFEHK--LQKAQDELDI
PARAPLEGIN IAVVMSLLNALSTSGGSIWNDFVBEMLAKCEVQRVQV PESDVVEVYLHPGAVVFCRPRRLALMYRMQVANI---DKFEK--LRAAEDELNI
YME1_YEAST PQYVVSRFETFGIASSPECHELYNEALQRIGRSEADAQRN-LLTASSAG-----AVNPSLASSSSNQ--SGYEGN--FFSMYSFLYG
SPAST_HUMAN IVTGQSGQECERARRLQARMHTFLVMAKDRLLQLEKMQP LPPFSKSQTDVYNOSTN--LACNGLQSESGAVPKRKDPLTHTSLSLPSKTVMTK

AFG3_YEAST PFDRIPIKVIERSGPFITLFPF---LPTIILLGCLYFITRKINSSPPNANGGGCGGLCGMFNVGKSEAKLENKETDIK---ISKNVA-CDE
RCAL_YEAST KDFRIPVLVQEGHNAKAMFOI---LPTAVGLIAGIWLTRRS-----AQAAGGSRGGIFGLSRSAKKNFTETVK---IKKNVA-CDE
PARAPLEGIN EAKDRIPVSYKRTGFCNALYSV---GNTAVGLAILWYVFRLAG-----MTGREGGFSAFNOLKMAFTIYDGMKMGK---VSKNVA-MHE
YME1_YEAST SRKE--PLHVVVSESTFTTVSRW---VKMLLVFGILTYFSFEG-----FKYITENTLLKSSEVADKSVDAKTN---VKDVC-CDE
SPAST_HUMAN GSAGLSCGHRAPSYSGLSMVSGVQSGGPAFTTRKGTPTKTRNTPSTPTTATRKKDLKNFRNVDNLANLIMNEIVDNGTAVKDDIAQDL

AFG3_YEAST KQIMFPHFLKN--GKYTKLGAKIFPAIS---TTLKNTGK---ANVP--LVVSGSEFVEMFV--VASRV--D--TQ--SMA--S--IF--I
RCAL_YEAST KKEIMFSELKE--SRYEKMGAKIFPAIS---TTLKNTGK---ACVP--YVVSSEFVEMFV--VASRV--D--KT--SMA--S--VF--I
PARAPLEGIN KLEVRFDYLS--EFLQLGAKVPAAL--L--C--L--K--V--T--AQVF--LAMAGPEFVEVIG--LAARV--S--KE--SMA--C--VY--I
YME1_YEAST KAELEIDFLKD--TKYESLGCKLPKVL--T--L--R--R--G--AGVD--PFMSGSEFDEVYV--VAKRD--D--AQ--SMA--A--IF--L
SPAST_HUMAN KQALOILILPSLEELFTGLRAPA--ELLF---N--M--K--V--A--SNAT--FNISAASLTSKYV--EKLVA--AV--ELQ--S--IF--V

AFG3_YEAST AIGKEGKGGALGGANDEREATLNQLVNM--FTTS--DQVVVLG--RPDV--N--LM--PG--FDRHIQIDS--DVNG--QIYLVHLKRLNLDPLLT
RCAL_YEAST AICKAKQK-HFSGANDERENTLNQVNM--FTPA--DQVVVLG--RPDI--K--LL--PG--FDRHINIDK--ELEG--KAIFAVHLHLKLAGEIF
PARAPLEGIN AVGKKSTT-MSGFSNTEEEQTLNQLVNM--MGTT--DQVVVLG--RADI--G--LM--PG--LDRHVFDIL--TLOE--REIFEQLKSLKLTQST
YME1_YEAST AIGCKNPK---DQAYAKOTLNQLVNL--FSQT--SGIIGAG--FPEAK--KLT--PG--FDKVVNVDL--DVRG--ADILKHHMKITLADNVD
SPAST_HUMAN SLLCE--REG---EHDASRLKTEF--IF--VQSAGDDRVLVMA--RPQE--E--VL--FIRRVVSL--NEET--LLLKLHLCKQGSPLTQK

AFG3_YEAST DDMNLSGKL--TL--PFTTADIANACNE--LIAARHNDPYITIHHFEQ--IERVIAGLEKKTRV--SKEEKRSVAYHEAGH--VCGWFLKYADPLKV
RCAL_YEAST DLKN---RL--AL--PFS--ADIANVCHNE--LIAARSDDAVKNHFEQ--IERVIGGVVERKSL--SPEEKVVAYHEAGH--VCGWFLKYADPLKV
PARAPLEGIN FYSQ---RL--EL--PFS--ADIANICHNE--LIAARECHTSVHTLNFEY--VERVLGATAKSKV--SKEEQKVVAHEAGH--LVGWMLEHTEAVMKV
YME1_YEAST PTII---RG--PLS--AELANLVNQ--VYACQKNVSVDMSHFEY--KDKILMGAERKTMV--TDAARKATAFHEAGH--IMAKYTNGATPLYKA
SPAST_HUMAN ELAQ---L--RM--DYS--SDLTALAKD--LGPIRELKPEQVKN--MS--SEMRNIRLSDFTES--KKIKR--SVSPOTLE--YIRWNKDFGDTTV--

AFG3_YEAST SIIPRCQCALGYAQYLPFDQYLISEEQFRHMINALGGRVSEELHFP--SVTSGAHDHDFKKVTOMANANVTSLGMSPKIGILSFDQNDG--NFKV
RCAL_YEAST SIIPRCQCALGYAQYLPFDIFLLTEQQLDRMTSLGGRVSEELHFP--SVTSGASDDFKKVTSMATANVTSLGMSDKIGWVYQKR--DSDL
PARAPLEGIN SITPRNAALGFAQMLPRDQHLFTKEQLFERMCMALGGRASEALSFN--EVTSGAODDLKRVTRIAYSNVKQFGMAPGIGFISFPEAQEGLMGIG
YME1_YEAST TILPRGALGITFQLPEMDKVDITKRECQARLDVCMGGRKIAEELIYCKDNTSSCGSDLQBATGTARAVTQYGMSSDVGFPVNLSENW-----
SPAST_HUMAN -----

AFG3_YEAST NKPFNSKARTIDLEVKSIVDDAHRACTELLTKNLDKVDLVAKELLRKEAITREDMIRLLGPRP-FREBNEAFKYLK-----PKSNTPEP
RCAL_YEAST TKPFSDETGDIIDSEVYRIVQECBDRCTKLLKKAEDVEKIAQVLLKKEVLTREDMIDLLGKRP-FPEMNDADFQYLNIDYETEKIRKEEKEKRR
PARAPLEGIN RPPFSQGLQOMMDHEARLLVAKAYHTEKVLQDHLKQALANALLEKEVINYEDIEALIGPPPBGPKKMIAPQRWIDAQR--RKQDLGEEETEET
YME1_YEAST -ESWSNKIRDIADNEVIELKDSERARRLTKKNVELBRLAQGLIEYETLDAHEIEQVCKGEK-LDKLKTSTNTTVVEGPD--S-DERKDIGDDK
SPAST_HUMAN -----

AFG3_YEAST PEAPAAATN-----
RCAL_YEAST NEPKPSTN-----
PARAPLEGIN QQPFLLGGEETWPK
YME1_YEAST PKIPTNLNA-----
SPAST_HUMAN -----

```

FIGURE 4B



6/7

Human: 1 [...] 459  
 Mouse: 1 AGGCCGAGAGCGTCCGCGTCTTCCACAAGCAGGCCCTCGAGTACATCTCCATTGCCCTGC 60  
 Human: 460 AGGCCGAGCGCGTCCGAGTCTTCCACAAGCAGGCCCTCGAGTACATCTCCATTGCCCTGC 519  
 Mouse: 61 GCATCGACGAGGAAGAGAAAGCAGGACAGAAAGGAACAAGCTGTGGAATGGTATAAGAAAG 120  
 Human: 520 GCATCGATGAGGATGAGAAAGCAGGACAGAAAGGAGCAAGCTGTGGAATGGTATAAGAAAG 579  
 Mouse: 121 GTATCGAAGAACTGGAAAAAGGAATCGCTGTTATAGTTACGGGCCAAGGTGAACAGTATG 180  
 Human: 580 GTATTGAAGAACTGGAAAAAGGAATAGCTGTTATAGTTACAGGACAAGGTGAACAGTGTG 639  
 Mouse: 181 AAAGAGCTAGACGCTTCAAGCCAAAATGATGACTAATTTAGTTATGGCCAAGGACCGTT 240  
 Human: 640 AAAGAGCTAGACGCTTCAAGCTAAAATGATGACTAATTTGGTTATGGCCAAGGACCGCT 699  
 Mouse: 241 TACAACCTCTAGAGAAGCTGCAACCAAGTTTGGCAATTTCCAAGTCACAGACGGAGCT 300  
 Human: 700 TACAACCTCTAGAGAAGATGCAACCAAGTTTGGCAATTTCCAAGTCACAAACGGAGCTCT 759  
 Mouse: 301 ATAACGAGAGTACTAACCTGACATGCGCAATGGACATCTCCAGTCAGAAAGTGGAGCAG 360  
 Human: 760 ATAATGACAGTACTAATCTGGCATGCCGCAATGGACATCTCCAGTCAGAAAGTGGAGCTG 819  
 Mouse: 361 TTCCGAAGAGGAAAGACCCCTTAACACATGCTAGTAATTCATTGCGCTCGATCAAAAAG 420  
 Human: 820 TTCCAAAAGAAAGACCCCTTAACACACACTAGTAATTCAGTGCCTCGTTCAAAAACAG 879  
 Mouse: 421 TCCTGAAAAGTGGCTCCGCAAGGCTCTCCGGTCAACACAGGGCGCTAGTTGGAGTGGTT 480  
 Human: 880 TTATGAAAACCTGGATCTGCAAGGCTTTCAGGCCACCATAGAGCACCTAGTTACAGTGGTT 939  
 Mouse: 481 TGTCCATGGTTTCTGGAGCAAGACCGGGAOCTGGTCTGCAGCTACCACACATAAGGGTA 540  
 Human: 940 TATCCATGGTTTCTGGAGTGAACAGGGATCTGGTCTGCTCTACCACTCATAAGGGTA 999  
 Mouse: 541 CTCGAAAACCAAAATAGAACAACCAACCTTCTACTCCCAACTGCAGTTCCGAAAAGA 600  
 Human: 1000 CTCGAAAACCAAAATAGGACAAATAAACCTTCTACCCCTACAACTGCTACTCGTAAGAAA 1059  
 Mouse: 601 AAGACTTGAAAATTTAGGAATGTGGACAGCAATCTTGCTAACCTTATAATGAATGAAA 660  
 Human: 1060 AAGACTTGAAAGATTTAGGAATGTGGACAGCAACCTTGCTAACCTTATAATGAATGAAA 1119  
 Mouse: 661 TTGTTGACAAATGGGACAGCTGTTAAGTTTGATGACATAGCGGGCAGGAGCTGGCAAGC 720  
 Human: 1120 TTGTTGACAAATGGAACAGCTGTTAAATTTGATGATATAGTGGTCAAGACTTGGCAAAAC 1179  
 Mouse: 721 AAGCGCTGAGGAGATTTGTCATCTCTCTGCGGCTGAGTTGTTACAGGGCTCA 780  
 Human: 1180 AAGCATTTGCAAGAAATTTGTTATTCTTCTCTGAGGCTGAGTTGTTACAGGGCTTA 1239  
 Mouse: 781 GAGCTCCTGCTAGAGGCTTGTACTCTTCGGTCCGCCAGGAAACGGAAAAACATGCTGG 840  
 Human: 1240 GAGCTCCTGCCAGAGGGCTGTTACTCTTTGGTCCACCTGGGAATGGGAAGACAAATGCTGG 1299  
 Mouse: 841 CTAAGCAGTAGCTGCAGAGTCTAATGCGACCTTTTCAACATAAGTGTGCCAGTTTAA 900  
 Human: 1300 CTAAGCAGTAGCTGCAGAAATCGAATGCAACCTTCTTTAATATAAGTGTGCCAGTTTAA 1359  
 Mouse: 901 CTTCAAAATATGTTGGGAGAGGAGAGAAATTTGGTGAGAGCTCTCTTTGCTGTGGCTCGAG 960  
 Human: 1360 CTTCAAAATACGTGGGAGAGGAGAGAAATTTGGTGAGGGCTCTTTTGTGTGGCTCGAG 1419  
 Mouse: 961 AACTTCAACCATCTATAATTTTATAGATGAAGTTGACAGTCTTTGTGTGAGAGACGGG 1020  
 Human: 1420 AACTTCAACCTTCTATAATTTTATAGATGAAGTTGATAGCCTTTTGTGTGAAAGAGAG 1479  
 Mouse: 1021 AAGGGGAGCAGCAGCTAGCAGACGGCTAAGACGGAATTTTAAATAGAAATTTGACGGGG 1080  
 Human: 1480 AAGGGGAGCAGCAGTGTAGTAGACGCTAAAAAAGTGAATTTCTAATAGAAATTTGATGGT 1539  
 Mouse: 1081 TGCAATCTGCTGGAGATGACAGAGTACTTGTAAATGGGTGCAACTAACAGGCCCAAGAGC 1140  
 Human: 1540 TACAGTCTGCTGGAGATGACAGAGTACTTGTAAATGGGTGCAACTAATAGGCCCAAGAGC 1599  
 Mouse: 1141 TTGATGAAGCTGTTCTCAGGCGTTTCATTAAACGGGTATATGTGCTCTTACCAATGAGG 1200  
 Human: 1600 TTGATGAGGCTGTTCTCAGGCGTTTCATCAACGGGTATATGTGCTTTACCAATGAGG 1659  
 Mouse: 1201 AGACAAGACTCCTTCTGCTTAAAAAAGCTGTTGTGTAACAAGGAAGTCCACTGACCCAAA 1260  
 Human: 1660 AGACAAGACTACTTTTGTCTAAAAATCTGTTATGTAAACAAGGAAGTCCATTGACCCAAA 1719  
 Mouse: 1261 AAGAACTCGCACAGCTTGCTAGAATGACCGATGGATACTCTGGAAGTGATCTGACCGCTT 1320  
 Human: 1720 AAGAACTAGCACAACCTTGCTAGAATGACTGATGGATACTCAGGAAGTGACCTAACAGCTT 1779  
 Mouse: 1321 TGGCCAAGGATGACGCGCTGGGTCTATCCGAGAACTGAAGCCAGAGCAGGTGAAGAATA 1380  
 Human: 1780 TGGCAAAAGATGACAGCAGCTGGGTCTATCCGAGAACTAAAACGAGAACAGGTGAAGAATA 1839  
 Mouse: 1381 TGTCTGCCAGTGAGATGAGAAATATTCGATTATCTGACTTCACAGAAATCTTAAAAAAGA 1440  
 Human: 1840 TGTCTGCCAGTGAGATGAGAAATATTCGATTATCTGACTTCAGTGAATCCTTGAAGAAAA 1899  
 Mouse: 1441 TAAACGCGAGTGTGAGTCTCAGACCTTAGAAGCATACATACGCTGGAACAAGGATTTTG 1500  
 Human: 1900 TAAACGCGAGGTCAGCCCTCAAACCTTAGAAGCGTACATACGTTGGAACAAGGACTTTG 1959  
 Mouse: 1501 GAGACACCAGTGTTTAAGGAAT 1523  
 Human: 1960 GAGATACCAGTGTTTAAGGAAT 1982  
 Human: 1983 [...] 3263  
 Mouse: 1524 GGATGCTCTGTGAGCCCATAGAACATCGCACTTCACAGGAAACAAGAGCTTTGGCTACA 1583  
 1584 GGAACCCAGACTTCGTTTACAGGACGTTTATAGAGTTTTCATTTTGTGCAACCAACTTGA 1643  
 1644 AGAGGAACAAGAAGACAGACCTAATAAATATGCAATATGAATGG 1689

FIGURE 5



7/7

FIGURE 6B

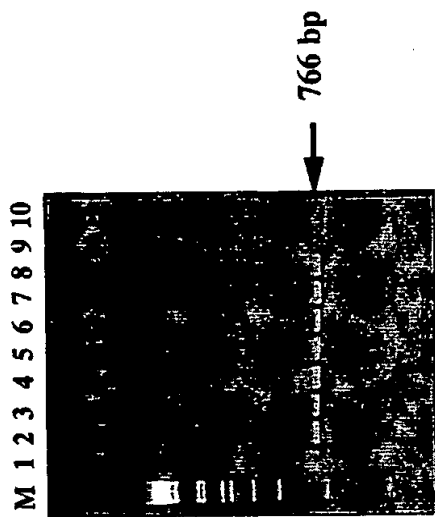


FIGURE 6C

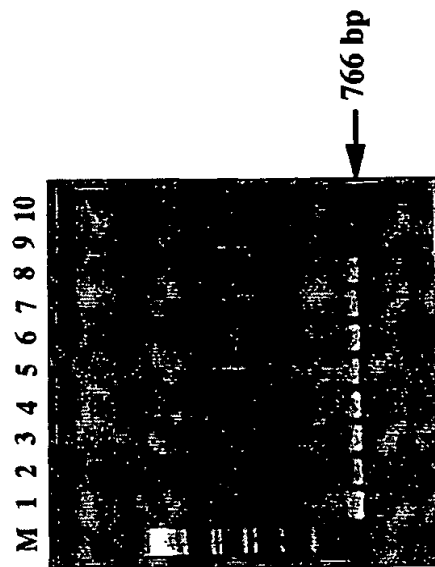
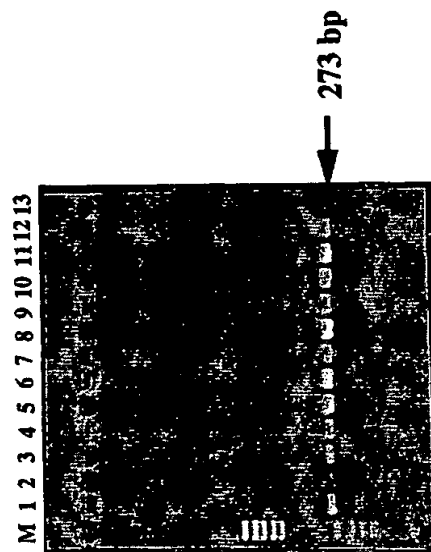


FIGURE 6A





## LISTE DE SÉQUENCES

<110> CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE - CNRS

<120> CLONAGE, EXPRESSION ET CARACTÉRISATION DU GÈNE SPG4  
RESPONSABLE DE LA FORME LA PLUS FRÉQUENTE DE PARAPLÉGIE  
SPASTIQUE AUTOSOMIQUE DOMINANTE

<130> D18374

<150> FR 99 11097

<151> 1999-09-03

<160> 107

<170> PatentIn Vers. 2.0

<210> 1

<211> 110000

<212> DNA

<213> Homo sapiens

<220>

<221> exon

<222> (9932)...(10471)

<220>

<221> intron

<222> (10472)...(33718)

<220>

<221> exon

<222> (33719)...(33805)

<220>

<221> intron

<222> (33806)...(35748)

<220>

<221> exon

<222> (35749)...(35832)

<220>

<221> intron

<222> (35833)...(45022)

<220>

<221> exon

<222> (45023)...(45118)

<220>

<221> intron

<222> (45119)...(60863)

<220>

<221> exon

<222> (60864)...(61051)

<220>

<221> intron

<222> (61052)...(61927)

<220>

<221> exon

<222> (61928)...(62061)

<220>

<221> intron

<222> (62062)...(62344)

<220>

<221> exon



<222> (62345)...(62438)  
<220>  
<221> intron  
<222> (62439)...(73173)  
<220>  
<221> exon  
<222> (73174)...(73248)  
<220>  
<221> intron  
<222> (73249)...(74633)  
<220>  
<221> exon  
<222> (74634)...(74705)  
<220>  
<221> intron  
<222> (74706)...(82788)  
<220>  
<221> exon  
<222> (82789)...(82864)  
<220>  
<221> intron  
<222> (82865)...(83102)  
<220>  
<221> exon  
<222> (83103)...(83194)  
<220>  
<221> intron  
<222> (83195)...(83334)  
<220>  
<221> exon  
<222> (83335)...(83414)  
<220>  
<221> intron  
<222> (83415)...(88129)  
<220>  
<221> exon  
<222> (88130)...(88172)  
<220>  
<221> intron  
<222> (88173)...(89561)  
<220>  
<221> exon  
<222> (89562)...(89641)  
<220>  
<221> intron  
<222> (89642)...(91162)  
<220>  
<221> exon  
<222> (91163)...(91233)  
<220>  
<221> intron  
<222> (91234)...(93443)  
<220>  
<221> exon  
<222> (93444)...(93484)  
<220>  
<221> intron  
<222> (93485)...(100599)  
<220>



&lt;221&gt; exon

&lt;222&gt; (100600)...(102009)

&lt;400&gt; 1

taactgctc	tgctgtgttt	ccttggcatt	atagctaate	aaattgagca	ggcaggttaa	60
cagtttaac	ttacacctac	tatttcaaaa	ccatgagctc	attcacattt	tcactgaagt	120
aacaaatcct	ccataaacta	gaaaatctca	aactgggtgac	tgggagtttt	ggttttgttt	180
ttttgttggt	ttattttatt	ttattttatt	ttctagatgg	agtcttgctc	tgttgcccag	240
gctggaatgc	aatggcatga	tctcagctca	ctgcaacctc	cacctttcgg	gttcaagcaa	300
tcctcctgcc	tcaatcttcc	aagtagctgg	gactacagga	atgagctgcc	gcacctggcc	360
tgggttgttt	gtttttaaat	tttgaggcca	ggtgcagtgg	cccatatctg	tgatcccagc	420
actttgggag	accaaggcag	gccgattact	tgaggtcagg	agttcaagac	cagccaggcc	480
aacatggtaa	aacatgtcgc	ctactaaaaa	cacaaaaatt	aggccgggca	cggtggctca	540
cgtctataat	cccagcactt	tgggaggcca	aggtgagtgg	atcacctgag	gtcaggaggt	600
caagaccagc	ctggcaaaaca	tggtgaaaacc	ccgtctctat	taaaaataca	aaaattagcc	660
gggtgtggtg	gcgggcacct	gtaatcccag	ctattgggga	ggctgaggca	ggagaatcgc	720
ttgaacctgg	gaggtggagg	ttgcagttag	tcgagatggt	gccactgacc	tccagcctgg	780
gcaataagag	cgagacgcca	tctcaaaaat	atataaataa	ataaataaaa	atacaaaaag	840
tagctaggct	ttgtggcagg	cacctgtagt	cccagctact	taggaggctg	aggcaggaga	900
attgcttgaa	cccaggaggc	ggagggttgca	gtgagccaag	attgtgccac	tgcgttccag	960
cctgggtgac	agagcaagac	tccttctcaa	aaataaataa	ataaataaat	aaataaataa	1020
ataaataaaa	ttttgagctg	ggcatgaaag	ctgaggcagg	aggatccctt	gagcccagca	1080
gtttgagacc	ccagttagct	ataattctga	cactgcactt	cagcctggct	gacagaggga	1140
gaccgtgtat	ctaaaaagaa	taaaataaca	atgatttttg	agccaataac	tcttagccaa	1200
tagatttccac	tataaaaattt	agaattcttg	tttctcttga	aaaattaaaa	aaatctgaca	1260
atgttaagct	tcacattcct	gaaaggcaaa	aatcagtggg	agctagctgg	gtgctgtggc	1320
tcacgcctat	aaaccagtg	ctttgggagg	ctagggtgaa	tggattggtt	gagtccagga	1380
gtttgagatc	agcctggcca	acacagtcaa	accccatctc	tacaaaacat	acaaaaatta	1440
gctgggatat	ggtggtgtgc	acttgtgtgc	ccaactactc	aggaggctga	ggcaggagga	1500
ccacttgagc	cctggagggt	gagggggcag	tgagccatga	ttgtgccact	gcactccagc	1560
ctggacgaca	gagagacatt	gtctcaaaaa	gaaaaaaaaa	aatcagctca	gtgggagctg	1620
agtaacagct	gtcctgttca	attacaggat	gcaactcttt	agcttctcat	agtttccatc	1680
attcacatca	tacatgcatt	tttgcctgcc	acacaaccct	cacatggaac	ccatatatgt	1740
tatatgcctg	accactgttg	ctattggaag	ttttggccac	tgcattaaac	tataaactcc	1800
atcttattaa	tcctgacacc	ccacttattg	cctgatatat	tgtccgtttc	ttaatatcta	1860
ttcatagaac	aaatgaatga	ataatatgtg	ccacattgtg	gactcaattc	aggagatga	1920
ctaatacttc	acaattatgt	ttttccttct	taatacagag	attggataat	aattccccaa	1980
ttgttacttc	tctcatccct	cccctccaac	catatctttt	catttgtttt	tcttatttta	2040
ggttttctgc	ctttttcaaa	atcagccatt	tcctcactgg	actctacatg	tgccattttt	2100
tttttttttt	ttttactaat	ttttttagtt	gaaaagaggt	ccttaatatc	tgtcattggt	2160
ccacacttga	aatctagaat	ctctcgaatt	gaaggtctga	agagttcctt	agaaggcaag	2220
tagggccgaa	atatccaaaa	aaaactattt	gatggtaggc	actgtggtaa	ataaatatgg	2280
gttaaagtga	taaagtgacc	gaatgtcctg	gattagttag	cacagtacct	agctccttct	2340
ttttgtcctt	gtgaaactgt	ctcagattcc	attcaagatt	aagtgtcctg	aaagttctta	2400
caagctgaaa	actgggaggc	tctcacactg	taggtagaat	ggctagcagg	gttgggatct	2460
catctaggca	cattgcagaa	ccagggtaac	accaaggcat	tatttttttt	ctttctttac	2520
ttttcttctt	tttttttaga	gacagggtct	cattatgtta	cccaggctgg	attccaattc	2580
ctgagctcaa	gtgatcctcc	tgagtagctg	ggacttcagg	tgtgcaccgc	tgtgtctgaa	2640
ggcactgttt	ttttgttttt	ttttgttttt	tttgacacaa	atttaggaag	atgttaattc	2700
acaacagtct	tgagactgag	atataattcc	aaggagcaga	agatgtgagt	ttagagcact	2760
aaaaaatgag	attccagtag	agtcagaaat	ctgaaatggc	attacagata	taagagacaa	2820
aaacaaaatg	tattgagctc	tgtcatgttg	caggcatcat	gatggaggtt	ttagatgtac	2880
tccttccattt	tgtaaattttt	atagagggaat	taactagaat	agcaacccca	gtcccaactaa	2940
ctctaaatcc	catatttttta	ccatacaaaa	aaagagcaaa	agtgcagaaa	agcacagtca	3000
atattaagta	caaccagata	gcagagactc	agtaaattggg	aggccggagg	cccgaacccc	3060
aacataatgc	cctatgaatga	aagcccatca	cttgcgcttc	aggggctaac	aatatactta	3120
tttcataatt	aaaatagaaa	cagactatgt	aaaaatatta	ttcttgagat	cccagatttt	3180
taccttaaat	tactgaggca	gtaagtgtaa	ttactaataa	tgtgatgttg	ggcaataaac	3240
agactcttag	agccccaat	tctttatttt	aaaaaactga	ggccagatga	ggtggcttac	3300



acatgtaatc	ccagcactgg	ggatgccaaa	gcaggcagat	cacttgagct	tgagcttagg	3360
agtttgagat	cagcctgggc	aacacggtta	agacctcatc	tctaaacaaa	atacaaaaat	3420
tagccaaaca	tggttggttg	cacctgtgtt	cctggctact	caggaggctg	aggtggaagg	3480
atcgcttgag	cccagagagg	tgagggtgca	gtgagccatg	atcttgccat	tgcaactctac	3540
actctagtct	gggtgacaga	gtgagactct	gtcttggggg	aaaacaaaag	agatgataat	3600
gcttaactga	agtagcaata	ttttaaaaag	gcactaaaag	ttcatctgct	tagttcagaa	3660
tatgggttcg	ataaatatta	gcaagtagta	gtagtcatca	tcatcactgt	cactgctgtt	3720
ctctccttaa	acttaagcat	gttttttgtt	tttttgagac	agtgtctcac	tctgtcacc	3780
aagctggagt	tcagtgggtg	gatcttggct	cactccaacc	tctgcctccc	aggttcaagt	3840
gattctcctg	tctcaccctc	ccaagtagct	gggaccacag	acacgtacca	caaccacacc	3900
cagctaattt	ttcgtatttt	tggttagagg	ggggtttcac	catgttggcc	aggctggtct	3960
tgaactcctg	acctcaagtg	atccacctgt	ctcgccctcc	caaagtgtg	ggattacagg	4020
cgttagccac	agcatccagc	cttaagcatg	ttaattaagt	ttttataatt	cagcaaaatg	4080
gttggaaaa	gctgtcttaa	atgagatgct	taagctgccg	tctgaacatg	aggtagaagg	4140
aaattctaca	cataatcatt	gtgctaaatt	acttgcaaa	atggccacaa	caattcctcc	4200
tatcctcata	tatatgcccc	tttgcaatgt	gactttgcta	cttctctatc	aagatgtgga	4260
gcttattttc	ccatatattg	cactagagtt	ggccttctga	cttgctttga	caatggaatg	4320
tagtacaat	gacactgtgc	aactttggat	tttaggtttc	gagagaactt	acaccttcca	4380
ctcacactct	cttggaacc	agatgcaatg	taaagaagtc	agggctatcc	tgctagagac	4440
atatgtccca	gctaatagcc	acaatcaacc	tctgaacata	tgaatgaggc	tagctaggcc	4500
atacagccat	tcggtcaagc	catcagatga	ctacatccac	aggaatgatc	cacaggcaag	4560
gccatcagaa	gaaccatcca	gctgaactta	ccccaaattg	ctgagtcaca	aagttgtgtg	4620
taaataaatg	tctgtctatc	taagccagtg	agttttggag	tggtatatta	catagcatca	4680
gaaatctaag	acaatcatta	tggttgaaatc	atttttcaaa	tttctcatat	ttattaaatg	4740
agtaaccata	gcaaggtgtc	aggctggatg	caaaaagtga	ggcaaaatgt	ataaagtgtg	4800
accactgcct	tcagtaagtt	tacaatctat	atcaagaggt	gatgaagtgt	ttaaataatc	4860
atcctgcagg	gcaatatagt	ataagagcca	cagagtaaca	caaccatatt	gtcataacaa	4920
ctgaaaaaca	agatcatttc	tgctggaggt	gataatggaa	taatttatca	agaatataac	4980
agagctggac	gcggtggctc	acacctgtaa	tcccagcact	ttgggaggcc	aagggaaggtg	5040
gatcacaagg	tcaggagttc	gagaccatcc	tggttaacac	gatgaacccg	tctctactaa	5100
aaatacaaga	aattagccgg	gcgtgggtgg	acgcgtctgt	agtcgcagct	actcaggagg	5160
ctgaggcagg	agaaccactt	gaacatggga	agcagaggtt	gcattgagct	gagatcgtgc	5220
catggcactc	cagcctgggt	gacagagtga	gactcagtct	caaaaaaaaa	aaaaaaaaaaa	5280
aaaatataac	attagaggta	agtcttgaag	gactttgaca	gtggaagtag	gaggcgaggc	5340
cattctaaat	gaatgaaaaa	tgacaggaga	gtaattgtag	tcctggaaaa	gagcaaaagta	5400
ggtacagacc	aacagtctat	attagctaga	gtatagtga	agtgcagagg	aaatgtcggga	5460
gaaccattct	ttattcaaaa	actatcttcc	tcatggccaa	gcatagtggc	tcatgcctgt	5520
aatcccagca	ttttgggagg	tcaaggtgag	tggtacactt	gagctcagga	attcaagacc	5580
atctggggca	acatagttag	acctcatctc	aactaaaaaa	caaaaaaatc	agacagatgc	5640
agtggtctac	acctgtatgc	ccagaacttt	ggtaggctga	ggcggggcga	tcacgaggtc	5700
aggagatcaa	gacctcctg	gacaacatgg	agaaacccca	tctctattaa	aaatacaaaa	5760
ttagctgggc	atggtggcac	atccctgtaa	tcccagctac	tcgggaggct	gaggcaggag	5820
aatcgcttga	accaggggagt	cggaggttgc	agtgcagcca	gatcgacca	ctgcactcca	5880
gtctggcgac	agagcgagac	tccatcttaa	aaaataaata	aattttaaaa	aaaactacc	5940
cagcatggtg	gtgcatgcct	gtagtcccag	ttactcagga	ggctgaggca	agaggggtgt	6000
ttgagccagg	gaggtcaagg	ctgcagttag	ctctgatggc	gccactgtac	tccagcttgg	6060
gtgacagagt	gagaccttgt	ctcaaaaaca	aaaacaaaaa	caaaaaacca	acaaatctcc	6120
ttgttagtat	catggtgagt	aaaaaataaa	ataaaaatag	aaataaactg	aacatgggtg	6180
ctcatgcctg	taatcctagc	actttggaag	gctgaagtgg	gaggattgct	tgagggctgg	6240
agttcaaaa	tggtctgggc	aacacggtga	gagagacctt	gtctctacaa	aagaactttt	6300
aaaacaaaaa	atagataatt	taaaaaaatt	aaaaaaaaca	aaazaataaaa	aaataatcaa	6360
gtatcaactt	gattccaggc	actgcttact	actctagtgt	tatactgtag	atgtggaagc	6420
tgagtaactc	atccaagatc	accgaaagtg	atggaacaca	gatctaaatg	caaccagtct	6480
gactccagga	ccatttaacc	attctactat	tgggccctat	cttggctaa	ttagaaagta	6540
agttactttc	tttagtggtg	aagactggag	ggataacagg	gaagatagtt	atttaagaaa	6600
aaaaactggc	atcaaaactaa	atatccatga	atagttgaac	agtaaaatag	ggttggtgta	6660
attcatataa	tggaataacta	tatagcatgt	aaaatgtacc	acagtatatg	aatctaacag	6720
ggaggaattt	caacacttaa	ttattaagta	ggtagccagg	catagcgggt	tatgcctgta	6780
atcccagcac	tttgggagac	caagacagga	ggattacttg	agcccagggg	ttcgagatca	6840



acctgggcaa	cagtgagact	ccatctctat	tttcttaaaa	taaaataaat	gaaatttttaa	6900
aaattttgag	gagggaaagc	aaacaaggga	tacttgaaat	atgattacat	ttccataaaag	6960
tcaaagtgag	gcaaaatcat	acaagacatt	gtttagaaat	acataaaatc	actgcaaact	7020
aaaaatgaga	cactagaatg	attaatataa	aattcaggat	agtggcttcc	tctagaggaa	7080
gagacaagac	attgagatta	gggaggagct	cacagagtgc	ttcgaggagt	tgggttacatt	7140
catttttctt	aaatggaatg	ctgcttatta	tttttcttta	aattgtgcat	ttaagtaaca	7200
cacttcttgt	ttatatgata	tatgtataaa	tgtaattttt	ttttttgaga	tggagtttcg	7260
ctcttggtgc	ccaggctgga	gtgcaatggc	actatcttgg	ctractgcaa	cctccacttc	7320
ctgggttcaa	gtgattctcc	tgccctcagcc	tcccagagtag	ctgggattac	aggcatgcgc	7380
caccatgccc	ggctagtttt	gtatttttaa	tagagaaagg	gtttctccat	gttggtcagg	7440
ctggtctcga	actcccagacc	tcagggtgatc	cgctgcctt	ggcctcccaa	agtgttggga	7500
ttacagggtg	gagccaccgt	gccaggccct	gaatcagatt	taaaagaggg	catttcatta	7560
aaaaaaattt	tttgttggtt	gcttttgaga	cagagtctcg	ctctgtcgcc	caggctgcag	7620
tgcatgggca	tgatcttggc	tcaccgcggc	ctcagcctcc	cagggtcaag	tgattctcct	7680
gcctcagcct	cgactagttt	gagattacag	gaatgcacca	ccaccacagg	aatgcacctg	7740
tctaactttt	gtatttttag	tatagaggga	gttttgccat	gttagccagg	ctgctcttga	7800
actcctgacc	tccggtgatc	tgctcgcctc	ggctcccaa	gttctgggat	tacaggcgtg	7860
agccaccaca	cccggccgaa	agagggcatt	tcagaatgag	ggctctagcat	aagcacagag	7920
aagggggagc	aataagaggg	aaacagggag	taggtcattt	ttgcaatagc	ctgtgacatt	7980
tgtagggcag	tactggcggg	gaataattaa	gtaaaatttg	ctggtgctgt	ggctcatgcc	8040
tgtaatccca	gcactttggg	aggccgaggg	gggcaggttg	cttgagccca	ggaattcaag	8100
accaacctgg	gaaacctagc	aagacctgtt	ctcaacaaaa	aagtaaaaaa	attagctggg	8160
ggcgcgatgg	gggtggctcat	gcctgtaatc	ccaacacttt	ggaaggctga	ggcaggcgga	8220
ttgcttgagc	ccaggagtgt	gagaccagcc	tgggcaacat	ggtgaaaacc	tggctctata	8280
aagaatacaa	aaattagtcg	ggcccagtg	cggtgtgcct	tgatcccagc	tactcgggag	8340
gctgaggtgg	aaggatcacc	tgagccaggg	agggtggaggt	tgacgtgagt	catgttgttt	8400
gcgccactgc	actccagcct	gggcaatgga	gtgaaaccct	gtccaaaaaa	taaaaaaata	8460
aagctgtggc	agaatgtgga	gattcttgga	agctggaagc	tctcatgggg	catttggaata	8520
cctcacattg	taaataacgg	agtcttttta	tcagtttggt	ttccttagtt	ttaggaaaca	8580
agaaataatt	atggctaact	caagtaaaaa	gagaaagaga	agagaaaaaa	gacgtggaga	8640
tagagagaga	gggagagaga	ggaaaagacg	aaaggaagga	agggggaaag	gagagaggaa	8700
gagagaaaca	gagaaacaga	ctgatttagt	tattggatag	attacataac	caagtgacca	8760
gtcaggaacc	cagcagctct	gggggagctc	aatgtgatgc	attgataaac	ccgctcttaa	8820
gagcactcgt	ttccagttac	tttctattcg	gtgggtctcc	agccaagatt	ccagggtccca	8880
ggagaatctg	actgacctag	tgtttgcttc	cgcttttgcg	gtctgggttc	tgtgcttgca	8940
gctcattaga	atacaggggag	cagagacaag	caggtagttt	cccaaaggaa	gggatgctga	9000
gtagattaaa	aaaaaagtgt	agattcttca	gtaaaactat	ggatggtaac	tatgcaaaac	9060
ctaagatttc	ccttattcaa	ataaattatc	tttcatatta	gacatctaaa	tatgactaa	9120
tttagttaaa	cccttgggtt	agttgatctc	atcacactga	gctaacattt	ttgttgctgt	9180
tgtttgtagt	gacctgaagt	ttcttatctt	cacaattgct	ttcctctcaa	ataattccca	9240
gatttttaaat	ttttatttta	ttttttctgg	agacggagtc	tcgctctgtc	gcccaggctg	9300
gagtgcagtg	gcgcgatctc	agctcacttg	cagcctctgc	ctcccagatt	caagcgattc	9360
tccggcctta	gccttccaac	cagctgggac	tacaggcgcg	cgccccacg	cccggctaata	9420
ttaattccca	gattgatatc	cattgcttct	gagatggggc	aattatcctt	cggagaagac	9480
ttaggctgcc	tggcagaaaa	agatgaaaga	aatctaagaa	aacgacgaca	ctgagagagg	9540
agcctagcga	accagcagag	cgaccccaag	ccgcaattcc	cccttccgtg	gatcgattac	9600
gaaggcttcc	tggcaggagc	tctccagggc	tgccgacgtg	aguccgaactg	cacattggga	9660
actgtagttg	agtgggaaag	ccgagaggcg	ggggccgcac	acgcgtacag	gggccccggt	9720
caacaaagac	gcgcggtgcg	cgcgcgcgcc	ggagaaaaac	acgggaagac	gtgcgcgtgc	9780
gcggccgccc	ctgggagcca	ccaggcgggc	gagaggacag	cgacagggaag	ggagggggccc	9840
gagccaccga	ctgcaggagg	agaaggggtt	gtgctcctgg	ccgagggaagg	agaaagggggc	9900
ggggccggcg	ggcagcgtgc	ggcagtgccg	agctcctgag	accggcgggc	acacgggggt	9960
ctgtggcccc	cgccgtagca	gtggctgccg	ccgtcgcttg	gttcccgtcg	gtctcgggga	10020
ggcgggttat	ggcggcgggc	gcagtgagag	ctgtgaatga	attctccggg	tggacgaggg	10080
aagaagaaag	gctccggcgg	cgccagcaac	ccggtgcctc	ccaggcctcc	gcccccttgc	10140
ctggcccccg	ccccctcccg	cgccggggcg	gcccctccgc	ccgagtcgcc	gcataagcgg	10200
aacctgtact	atttctccta	cccgtgtgtt	gtaggcttgc	cgctgctgcg	tttggctcgc	10260
ttccacctgg	ggctcctctt	cgtgtggctc	tgccagcgct	tctcccgcgc	cctcatggca	10320
gccaaagagga	gctccggggc	cgcgccagca	cctgcctcgg	cctcggcccc	ggcgccgggtg	10380



ccgggcccgcg	aggccgagcg	cgtccgagtc	ttccacaaac	aggccttcga	gtacatctcc	10440
attgccctgc	gcacgatga	ggatgagaaa	ggtaactagg	gggctggggg	agggggcggc	10500
ggcgccggga	agaaggcggt	ggggtcgccg	ggggagggca	acacctgcgt	cccttttctg	10560
cgggaggggga	cggtgcaccc	ccggaattga	tatgccccgg	gagactgctt	ccccgtaggt	10620
cgagagcctca	tcttctagta	ttcttaaaac	ctctccccct	tcagggcact	gtagctgtcg	10680
actttgtttc	agacaccagc	cttccccac	acttctgcat	gacccaggtc	actatgagac	10740
accagacgt	gttgatgaca	gtgacatttg	tcctagagtg	accacactga	tcctttctag	10800
cactgtgaag	agtgtgcagc	ttcctctgaa	ccaaggtttc	caaaaggttt	tgatattgaa	10860
raagcagtg	cgccttactg	gcttttaatg	aaagcagagt	attgtagtgt	cagaaaaaaa	10920
acaaatgg	tgacaatttt	gaaagaaata	gctgcatatg	actgcagttg	aatttgcac	10980
tttaatca	agataatcat	tttatcagta	caacgattcc	tgaatacttt	ttcaatgaag	11040
tttag	cataaaa	ttccccctgt	tgctttgatt	ttaattaaaa	cagtattcca	11100
ttctt	taattttccaa	gttgaaatgt	ttgatgaatg	gattgcgtaa	acttaacat	11160
ac	agtaaaacct	aaaacaacta	tgtatgtttc	tgaatgaaag	caaggatact	11220
	ggtttctttt	aagctaactt	tttttttttt	ttgagacgga	atctcgccct	11280
	tgagtgca	atgggtgcagt	ctcggtccac	tgcaacctcc	gactccctgg	11340
	ttctgcct	cagcctcccg	agtagctggg	attacaggca	cgcaccacca	11400
	acttttgtat	tttttagtaga	gtagggattt	caccatgttg	gccaggatgg	11460
	ctgacctcgt	gatcggtg	cctcggcctc	ccaaagtgt	ggtaatacag	11520
	ccgcgcctgg	ccttaagcta	acatttttta	ttatatgtgc	caggcattgt	11580
gct	ttgctatgtg	tatctttttt	cttttttaaaa	aaaatagcaa	ccatcccaga	11640
gaaaag	tgttatttaa	tcctcacaat	aactttgtga	gatgaaggta	ttattggtat	11700
gcatttta	gagatgaaga	aaatgaggcc	caaaagataa	aggaggttat	tccaaacctg	11760
gctaataat	gaaatttctt	ttatggaaaa	taagtgaat	taggagaagt	ctaactttta	11820
caattctccc	cttttatcct	tacttccagt	atgctgagat	cttgcttctc	cctctgccaa	11880
aaacacccac	ttttctacca	cacctcaatt	agatactcac	ttgcattgtc	cattagttaa	11940
aacagaaaca	atctgcactt	cattcatagt	gtctgtctct	actgccaact	ccaaaaactg	12000
ttctagaatt	tcaatcttgt	tgaaacctgt	ttcctttgtg	gggcctggga	gtgggaagtg	12060
ggatataagg	aaagaggcag	ttattcattg	ttttggacag	taaggaaaga	gtgacggtta	12120
agagaggtca	aggagggtgt	taacatttaa	gaatactatg	tgttttaga	aggaaatttt	12180
ttgttaactg	ctccattctt	tttttttttt	ttttttgaga	tggagtctcc	ctgtgtcacc	12240
cagactggag	tgacgtggcg	tgatctcagc	ttactgcaac	ctctgcctcc	cagacgcagg	12300
cgattctctt	gcctcagtct	cccaggtagc	tggaaccaca	ggtgcgcctc	atcacggctg	12360
gctaattttt	tgtattttta	gtagagatgg	tgtttcacca	tgttggccag	actggtcacg	12420
aactcctgac	ctcaggtgat	tcgcctgcct	cgacctcca	agtgcctgga	ttacaggtgt	12480
cagccactgt	gcccacccag	caagctccat	tctttatcac	ctcttaagaa	catccaggat	12540
cccttgggga	gaattaaaac	ggttgcaaag	ttttagaata	gaggaacatg	tttaagcgta	12600
gatcattttt	tgggccagtc	atgggtggctc	acgcgtgtaa	tcccaacact	ttgggaggcc	12660
ggggcggtg	gatcatgagg	tcaagagttt	ggacacatcc	tggccaacat	ggtgaaattc	12720
catctctact	aaaaatacaa	aaattagctg	ggcatggtga	cacatgcctg	tagtcccagc	12780
tactcaggag	gctgaggcaa	gagaagtgtc	tgaacctgga	agggtggagg	tgacgtgagc	12840
cgagatcggt	ccactacact	ccagcctggg	cggcagagca	agactccgtc	ttgaaaaaaa	12900
aaagtaggtc	atttttggct	gggcacgggtg	gctcatgcct	gtaattccag	cactttggga	12960
ggctgaggtg	ggtggattgc	ttgagcccag	gagtttgaga	ccagcctggg	caacatagtg	13020
aaaccctgtc	tttgtgaaaa	atacaaagat	tagctaggcg	cagtggcaaa	tgctgtagt	13080
cccagctact	tggggggctg	aggtaggagg	atcacttgag	ctcaggttgt	tcaggctgca	13140
atgagctgag	atcgtgcaac	tgactccga	actgggtgac	aggagtaaaa	ctgtctcaaa	13200
aaaaaaaaaa	aaaaaaaaacca	aaaaaaaaaaaa	aactgtttta	attgttttat	ttaggaaag	13260
aaagtcagaa	catgcaagga	aatttttttt	atttgtttat	ttttgagacg	gagtctcgct	13320
cagttgccca	ggctggagtg	caatggatg	atcttggtct	actgcaacct	ctgcctccc	13380
gattcaagcg	attctcctgc	ctcagcctcc	tgagcagctg	ggattacagg	tgtatgccac	13440
cacgcccagc	taatttttgt	gttttttagta	gagatggggt	tccaccatgt	tggccaggct	13500
ggtttcgagc	tcctgacctc	aagtgaaccg	ccctccttgg	cctcccaaag	tgctgggatt	13560
acaggtgtga	gccgcggtga	ccgaccacaa	ggaaatttta	gttaacactg	ttggttgatg	13620
ggagttggga	ggtaggataa	aaggagaaat	taaggaaaac	ctaggcatga	aaaataaaag	13680
accctgagct	cttagatttg	aagaaatagc	agttccatgt	gaggaataag	tggaagaaat	13740
agaattcaga	cctcaggtct	caggctggtg	acttaaatct	ttcagtatca	catatatgaa	13800
tatatctatt	atacttaaat	cctcctaacc	attttttatt	ttcagttgga	tatatataat	13860
atataaaaa	aattatttaa	tttatttttag	agacagtggt	tcactctctt	acccaagctg	13920



gagtggaaatg	tgatcatagc	tcactgcagc	ctcaaggctc	attcctaggc	ttaagtgatc	13980
ctctttttttt	ttttttttgct	agagatagga	tcttggttatg	ttagccaggc	tggagaaatt	14040
tcaacaatat	tttgaacaat	aaaaaaaaaaa	taaattaggt	tttattgtaa	agtggatatgt	14100
ttaagttttac	gccattctcc	tgcgtaatgc	atatttcata	ctcttcctac	tgataatggt	14160
tctgttcaca	actttttctt	tatattttga	tttcttttct	cttttctttt	tttttttttt	14220
tgagacagag	cctcgctttg	tcgcccaggc	tagagtgcag	tggcgcgatc	tcggctcact	14280
gctagctccg	cctcccgggt	tcacgccatt	ctcctgcctc	agcttcccga	gtagctggga	14340
ctacaggcgc	ccgccaccat	gccagctaa	ttttttgtat	ttttagtaga	gacgggggtt	14400
caactgtgtta	gccaggatgg	tctcgatctc	ctgacctcgt	gatccaccgc	cctcggcctc	14460
ccaaagtgc	gggattacag	gcgtgagcca	ccgcgcccgg	cacatatttt	gatttctaatt	14520
tgggacatca	gaatgggctg	ttgacagctc	tttttttttt	ttttgaggca	gagtctcgct	14580
tcttcaccca	ggctggagt	caatggcgcg	atctcagctc	actgcaacct	ccacctccca	14640
ggttcaagcg	attctcctgc	ctcagcctcc	ccagtaactg	ggattacagg	catgtgccac	14700
cccacaccca	gccaattttt	gtatttttag	tagagacagc	atctcaccat	gttggccagg	14760
ctgttctcga	actcctggcc	tcaagtgatc	cactcgcctt	ggcctcacia	actgctagga	14820
ttacagggtg	gagccaccgt	gccagcctt	gacagggtct	tagtttgatt	ttagttcaac	14880
aactgatgcc	gtaatatgcc	aaattaaatt	agttcagact	gaaacggatt	acttaaagat	14940
tcattttcct	ttaaaaaatga	agtaaaactt	tagccggatg	tgggtggcggg	tgtgtgtaatt	15000
cccagctact	cgggagggtg	aggcaggaga	atcgcttgaa	cccaggagggt	ggagggtgca	15060
gtgagctgag	atctcaccca	ttgcactcct	gcctgggtga	gaagagtga	atcccatcaa	15120
aaaaaaaaaaa	aaaggctggg	catggtggct	caagccgggt	gcagtggccc	atgcctgtta	15180
tcccagcact	ttggggaggca	gaggcaggcg	gatcacttgt	agtcagaagt	tcgcgaccag	15240
cctggccaac	atattagtga	aaccccgtct	ctaataaaaa	tacaaaaatt	agctgggcat	15300
agtgtgcac	gcctgtagtc	ccgggtactt	gggacgctga	cgaggacaa	ttgcttgaa	15360
tccagagggtg	gagattgcgg	tgggccaaaga	tcacgccact	gcacgccagc	ctgggcaaca	15420
gagcaagact	ccatctctca	aaaaaaaaaaa	aaaaagaaaa	ctgagtttat	attgttatgg	15480
ttcttatcta	tcttaaat	tttcttttag	agattgaata	ttttgtact	ttagtcttaa	15540
tccagtgggt	gaaagagagt	gcacttttga	agtctgcctc	ttggctgtcc	ttgacaacac	15600
aaaccttagt	tccaagagaa	tgtaatcttc	cctctttctc	agtgttcaa	aatatatatg	15660
atcaaataga	aacggagtga	ataggcagtc	tcttcaaagg	tttcttaact	tattctgtaa	15720
ctatcaagga	gctggttaata	tcattgcactg	ccattccctt	ggcaacatga	cttatctttt	15780
tttttttttt	ttctttttttg	agacagagtc	ttgctctgtc	acccaggctg	gagtgcagtg	15840
gcgagatctc	agctcgctgc	aagccccgcc	tcccacgttc	acgccatttt	cctgcctcag	15900
cctccaagt	agctgggact	acaggcgcct	gccatcacac	ctgggctaatt	ttttttgtat	15960
tttttagtag	agacgggggt	tcaccgtgtt	agccagtacg	gtctcgatct	cctgtcctcg	16020
tgatccgccc	accttggcct	cccaaagtgt	tgggattaca	ggtgtgagcc	actgcgcctg	16080
gcctacttat	cttctaattt	aactgaaaac	caatttat	gattcagtga	aatggcatca	16140
aactgtagta	gtgttaattg	aaatatttgg	taccttgaaa	tgtaaattgc	caaattaaat	16200
ctcattttaa	tgaatctcgc	tgtaaatgtg	ttttatatgg	tgactatggt	tattctgaa	16260
tttattctta	tggcatacta	aaaaaaaaaaa	aaattttttt	ttttttttgt	aatggagtct	16320
tgctctgtca	cccacgctgg	agtgcagtg	cgcaatctcg	gttactgtga	acctccacct	16380
cccgggttca	agcgattctc	ctgcctcagt	ctcccagta	gctggtacta	caggcgtgca	16440
ccactatgcc	tggctaatat	ttttgtattt	ttaggagaga	cagggtttca	ccatgttgg	16500
caggctgatc	tcgaactcct	gacctgaagt	gatccgcctg	cctcagcctc	ccaaagtgt	16560
gggattacag	gtatgaccca	ctgcacccaa	cccatactca	aatgtgacac	tgaattttca	16620
taaaggcctt	aatttatctg	aaaccaaact	atttcaaaag	aggaatagca	cagcaaatc	16680
tgttgactta	atgagaggat	atgtgaagtc	tatttattaa	agcaaatatt	aatggaggc	16740
cagtttaattt	gtacagctct	gcatttttaga	tatttgagaa	atattttatt	cctctccagt	16800
gagatgtgtt	aaaacattag	ttatgtgatt	aacaaatatg	tgtacatacg	tatatatgta	16860
catacacatt	ttgagacagg	gccttggctc	gttggttcagg	ctggagtgca	atgacaccat	16920
cttagctcat	tgaagcttca	gcctcgcagg	ctcaatcgat	ccacccacct	cagcctccct	16980
agtagctggg	tctacagaca	tataccacca	tgcttggctg	attttttaatt	ttttttaga	17040
gatggtgatc	ttgccctgtt	gactaagttg	gtaaatattt	taattgttga	actttcttgg	17100
aagactgaaa	acctgtgata	gccattttat	ataaggagaa	gctgaagttc	aaagagtaga	17160
ctcatagcag	aaacaaaaat	agaatttaag	tgaatggact	caaaatattg	tactttttac	17220
tttatactgc	aggtttttat	gttgtaattg	tggtaatgag	ctccttggaa	tatttggagg	17280
aaaagactga	gttgtaataa	tggttctttg	gatttactag	aacatatcat	gttctgcatg	17340
gctcctgtgg	gtagacaagc	cggaaatctc	ctgggtaaca	caatgggtgga	ggttctctag	17400
gtgacgtttg	atcttctcaag	tacataggac	taaacagaaa	aggcctagta	tggtatatga	17460



atgagagatc	aagtttctca	ggatattcta	gggctaaagg	atcaggcatc	gaagacagaa	17520
attgtcctaaa	taaaattttt	tctattcata	gttttaagg	gctaaaggg	caggcattga	17580
agacagaaat	tgtctaagta	aaatattttt	ctgttcata	ttttaagctg	tgtatatgtg	17640
catgtgtgta	tttaaaataa	cttcgtcaat	gaaaagagtc	aaactctgta	aaatatttga	17700
agagatttat	tctgagccaa	atatgagtga	ccagtgggcc	atgacacagc	cccagtagat	17760
actaagaaca	tctgtccaag	gtgggcaggc	tatagcttga	ttttatacac	tttagggaga	17820
cataagacgt	cagttaaaca	tgtaatgt	acattggttc	catctggaaa	ggcaggaaaa	17880
ctagaagttg	gggaggcttc	caggctcgtg	gcagattcaa	agattttctg	attggcaatt	17940
ggtcaaaaaga	gcttatctaa	agtcctggaa	tccatagaag	ggagtgtctg	gtttaaaata	18000
ataggttgta	gctaccaagg	tttttattgt	acagatgaag	cctccaggta	gcaggcttca	18060
gagagaatac	attgtaaatg	tttcttatga	gactttaaaa	ggcggcagac	tcttaagtta	18120
attttctcct	ggttcaggta	aaagacttgg	aaagggaag	gattctctac	agaacgtaaa	18180
ttttcccccac	aagagaaagc	tttgcagggc	catttcagaa	tatgtcaaag	aaatataatt	18240
tagggtaaaa	tacttcaatt	tgttttattt	atttatttat	tttttgagac	agagtctcgc	18300
tctgttgccc	aggctggagt	gcagtgccat	gatctcggt	cactgcaagc	tccgcctcct	18360
gtgttcacgc	cgttctcctg	cctcagcctc	ctgagtagct	gggactacag	gcgctggcca	18420
ccatgcccag	ctattttttt	tgtattttta	gtagagacgg	gatttcacca	tggtggccag	18480
gatggtctcg	atctcttgac	ctcgtgattc	acccgcctcg	gcctcccaaa	gtgctgggat	18540
tacaggcgtg	agccactgtg	cccggcctca	aaatacttca	atttctttca	tggcctgcta	18600
tctgacgtga	tgctgtacta	gagtcaggct	gggaatttgg	cgtcttattg	ctacaaaaca	18660
tcttaatatc	tctgttttaa	tgtaaatgct	gatcagttgt	ccctgaattc	caaaggggaag	18720
agggatatatg	agtcactgtcc	aacccccact	tctcattatg	gcctgaacta	gttttttagg	18780
tttaactttgg	aatgcctttg	gcaaggggag	gggtccatgag	tcagttgggg	gtcttagagt	18840
tttattttttt	gcttaccggt	ttataaaaaag	ttaatgaaaa	ttatcatatt	tcataattct	18900
atataattca	atattgtact	ttatttaaaa	ctcacgtata	aaatagctgt	ccatatctgt	18960
tttcagaaga	tgaggatgga	gggtagaaat	cagaagtgtc	agatttggtg	attttcttac	19020
actgctgaaa	acctatacct	accactttga	aaggattaat	ttcagacttg	ctttcttttg	19080
gcctaattgat	tctactttga	agtttctctg	attaaactaa	ggaataaaatc	tgataaatgg	19140
acattcagat	gataccatac	ttttccaaaa	gataacattg	cttttgatta	catatgcaat	19200
aaacatttca	cattttttct	caagattatt	tactggcattc	tgaccaaag	acacaaaaaa	19260
gcagccactg	tttaaggactt	tatcccttgt	tctgttttta	gctggtttgt	tggtgtattt	19320
ttcctcatgt	tgaatacagt	taaaccctat	taaactggat	tcccatatt	actgttagtt	19380
gtcctgatga	caattaagga	tagttagtaa	atggatattg	aatcatttta	tttttttagt	19440
agtgatggcg	tttcgctgtg	ttggccagg	tggtctcgaa	ctcctggcct	caagtgtct	19500
gcctgtgttg	gcccctcaaa	gtgttaggat	tacaggcgtg	agccactgca	cctggctgga	19560
tattgaatta	ttgaagaagg	atgttcccta	acacttctctg	cctcttttcc	gctgtcttac	19620
tctctctgtt	ccaccagta	gtagttgggt	ttttgatgtc	tggaagaa	tagattgtct	19680
actcttctat	attagttttc	ttcactcttt	tttttaaaaa	tggttttaaa	gcaagataga	19740
gacagggtct	tgccgtattg	cccaggctgt	tcttgaactc	ctgagctcaa	gcaatcctcc	19800
cacctcaccc	cataaagtgc	ggagattaca	ggcatgagcc	accatgcctg	gccagttttc	19860
ttcactcttg	acctataata	gtcctgcaaa	gccagtgaag	ctgttaatat	gctgacgtag	19920
cctttttttc	tctcgtttgt	gaattattaa	caatcgctga	tcttacacat	catatacaat	19980
aaaacatctt	tgaacactgt	cacatctcag	cagctcattc	tggttaatga	ggaaagaaaa	20040
atgtcaaaat	ctgtgatttt	cttagagggt	attaaatgtt	ttacagctat	gtagatattc	20100
tgtagacttg	ctgtacttac	ttacataact	tttctgctct	tctgcagagg	gagagaatta	20160
acttcataag	tgggctttgt	caatgccctc	cagtcagcga	tctccaggaa	caaacttatc	20220
tatggttgag	caagtgggg	ttattaccca	ctgtagccag	ggagaacaca	catatgtaat	20280
aaccatggta	tgtcatagta	aagggtgtta	ggaagatagg	atttgggctc	gtttatttgg	20340
tgattttaag	gaggggttaa	agaagcaggg	ttttgctctg	atttggatgc	taccaggagc	20400
tgcaataaat	tctatgaata	aatatcataa	aacctatcta	gaagaaaaga	ctagagtga	20460
gtctaaagct	gtagtagtta	aaaagcaaag	gtcactcctt	atctggaaaa	ggggaatatt	20520
tggtatttca	tggttttagac	agtgttcagt	gttcatgttt	tgctgtgtt	tagacataat	20580
tgtagagtgg	tcttggtcca	ccgtggtcac	agagtggcat	tatttcatgc	tgatattctg	20640
taagtgccag	gtcaaatacc	aaggtttagc	tgatagtagt	aggccagctc	ctggatgtaa	20700
taagctgttt	ttctctttct	cagctttttt	ttttttgtgt	ttttttttta	aatagagtct	20760
ggctctgtca	cccaggctgg	agttcagtg	gggcaatctc	agctcactgc	agcctctgcc	20820
tcctgtattc	aagcgactct	cctgcctcag	cctcccaact	agctggaatt	acaggcgcaa	20880
gtcaccacac	ccggctaatt	tttgtatttt	tagtagagac	ggggtttcac	catgttgagc	20940
aggctggtct	cgaactccgg	acctcaagca	atccatctgc	ctcagcctcc	caaagtgtctg	21000



agattacagg	catgagtcac	tgcacctggc	ctgtttctta	gatttgaggg	tcaactttta	21060
cccttttcga	actgtggggc	ttcttatgga	aattgacatt	taagtcctga	ccatatagga	21120
tcttgggcaa	gttactttaac	tatgtatgct	gagatgtttt	tttaaagtgc	tagtgcttgc	21180
acataatagg	tactcagtaa	atgagaacta	ttattataaa	atcaatagta	cttttaagat	21240
tacagctaga	ggttatgtta	gaacattgtt	aactcttcgt	actagtttct	gttacacttt	21300
ttaggatgca	gtagcagcat	aaacgagata	tggggaagaa	cgagaggtat	ttaaacagt	21360
atataggccg	ggcgtgggtg	ctcacgcttg	taatcccagc	actttgggag	gctgaggtgg	21420
gcagattacg	aggtcaggag	attgagacca	tcctggctaa	cacggtgaaa	ccccatctct	21480
actaaaaata	caaaaaacaaa	attagccggg	cgtggtagtg	ggcgcctgta	atcccagcta	21540
tttgggaggc	tgaggcagga	gagtggcgag	aaccaggag	acagagcttg	cagtgagccg	21600
agatcgcgcc	actgcacccc	agcctggggc	acagagcgag	actctgtctc	aaaaaaaaaa	21660
aaaaaaaaaa	attaaaaacaa	aaatatttgt	gttaattgtg	atgacaaaaa	aaaaaaaaaga	21720
gatggaaagtc	tctccctaac	ctcactcctc	atttagtgtc	atggcttttt	tctttttttt	21780
tttttttttt	ttgagacaaa	gcctcactct	gtcacccagg	ctggactgca	gtggtgcaat	21840
ctcagctcac	tgagggtctc	gcctcccaga	ttcaagcaat	tctcctgcct	cagcctcaca	21900
agtagatggg	actacaggca	catggcacca	tgcccagcta	attttttttg	tagtttttag	21960
tagagacagg	gtttcactat	gttggccagg	ctggtctcga	actcctgagc	tcaagtgatc	22020
ccccctcctt	ggcctcccaa	agtgtctggg	ttacaggcgt	gagccctgct	cccagactcc	22080
tggctttttt	tttttttttt	aatgaaaaat	tcaaaatgct	cttttttttt	tttttttttt	22140
tttgagacag	agtctcgctc	tgtcgcccag	gctgaagtgc	agtgggtgta	tctcggctca	22200
ctgcaacctc	tgcctcccag	gttgaagtga	ttctcctgcc	tcagcctctc	gaacagctga	22260
gattacaggt	acgcgcacc	acgcctggct	aatttttgta	tttttaatag	agatggggtt	22320
ttgtcatgtt	gtccaggcta	gccttgaact	cctcgctcgt	tgtgatccac	ctgccttggc	22380
ctcccaaagt	gctaggatta	taggcgtgag	tcgctgcacc	tggccacaga	aatttttttga	22440
agaagataaa	taaggtgaca	tttttaagg	tcaaagaaaa	tgtcaaaaac	tagaatgatg	22500
tctttacata	gggtttaaaa	ctttccaaat	taacagggaa	aataattctt	taccttgaaa	22560
ataaatgttt	gctagtgaat	gcaaatacaa	tctttttact	aaatgtttta	ttaaattttt	22620
ttttcttgta	gacacaggg	cctactgtat	tacccaggct	ggtcttgaat	gcctggcctc	22680
aagcaatcct	cctacctcag	tcttccaaag	tgctgggact	acagacatga	gccatcacac	22740
tgggtctttt	ttaccaaat	atagtagaaa	gcactttttc	tctaattggtg	aactatgaga	22800
gaattaatca	ggggctatta	gtaattcatc	cctgaattaa	tcagtgatta	taatgtcttg	22860
tgggtccatgt	agtttgctgg	ggattaacac	accatgaaag	tctaccagga	gatttttttt	22920
tttctttgag	aacaggggcc	ataatcagta	gtccttaaat	gaaatggact	attcccat	22980
cattatatgt	tgctaggct	ggactcgagc	tccttggtt	aagtgatctt	cccacttcag	23040
actctcaagc	agctgggact	ataggtgtgt	gccaccatac	ctggcttaat	tgagtgtttg	23100
ttttttgttt	ttttttctga	gacagagtct	cgttgtgtcg	cccaggctag	agtgcagtgg	23160
cacaatctcg	gctcactgca	acctctgttt	cccaggttca	aacgattctc	ctgcctcagc	23220
ctcccaagta	gctgggatta	caggcgcctg	ccaccatgcc	tggtctaatt	ttgtattttt	23280
agttagatg	gggtttcacc	atcttgcca	ggctgatctc	gaactcttga	cctcatgatc	23340
cacacacctt	ggcctcccaa	agtgtggga	ttataggcgt	gagccaccgc	gcccggcttt	23400
aattgagatt	tttagatatc	tattactctg	ctaattttgt	cacttgcaag	ttgccatcag	23460
aaaattgtag	gaaaatggat	atatttgttc	cttggaaatg	tttgtgtgag	aatacttaag	23520
gattaaatag	ataagtaaaa	ctgggtgggt	ttatataaca	tagatgagca	aatgtcagga	23580
acatacaact	gtgcacacag	ttcaggagaa	ggaggattta	agttaatcaa	caaatttact	23640
aagtataata	aagatactaa	aagtagtggt	tccataccac	tttattactt	aaagtatcat	23700
catataccct	attttatatg	atttttgcca	caagtcagag	ttaggtaaaa	gaaatacttg	23760
cttttcagg	aaggagtttg	acgcccagac	agattaactg	acttttccaa	aatcatattg	23820
ctattaaatg	gtggaacaag	gacttaaatc	tttgccctct	aactcacata	cttgcaaaaca	23880
catatcctct	cactctaccc	caagctaccc	atgttttgac	ccttcttggtg	gcaatctggg	23940
tctcactaat	atttgaaaga	aaacgtacag	tagataaatt	gcaagttaat	ctgttacgca	24000
tatctcttac	ctctatttaa	agatgaatat	cagcatttct	gttgtttcta	cagtaacata	24060
ctaaaaaata	atgcagtcca	ggtgcagtgg	ctcacacgtg	taatcctagc	acttttggaa	24120
gttgaggcag	gaggatcact	tgaagccagg	agttcgagac	tagcctgggt	atgcaagacc	24180
ccatttctgt	tttttttttt	tgtttgtttt	gttttgtttt	gttttgtttt	tttgagacgg	24240
agtctcgctc	tgtcgcccag	gctggagtgc	agtggcggtg	tctcggtcca	ctgcaagctc	24300
cgccctcccg	gttcacgcca	ttctcctgcc	tcagcctccc	aagtagctgg	gactacaggc	24360
gcccgcactg	acgcccggct	aattttttgt	atttttagta	gagacgggtg	ttcaccgttt	24420
tagccgggat	ggctcgcgac	tcctgacctc	gtgatcccca	tttctttttc	attgcaacct	24480
ttgcctccca	ggttcaagtg	attgtcgtgc	ctcagcctcc	caagtagctt	gggattacag	24540



.

.

.

.

gcatgtgcc	ccatgcctgg	ctaatttttg	tatttttttg	tagagatgga	gtttctggcc	24600
aagctggcca	ggctgggtctc	gaacgcctag	cctcaagtga	tctaccctcc	ttagcctccc	24660
aaagtgtctg	gattacaggt	gtgagccact	gtgccgggcc	ccatttctac	aacaattaaa	24720
aaatattagc	ccagtgtagt	ggtgcatgtc	tctagtccca	gctactcaga	aggctgaagt	24780
gaaaggattg	cttgagccca	gaatttcaag	gctacagtga	gctatgataa	tggcattgca	24840
ctccagcttg	ggtgacagag	tgagaccctg	tctctaaaaa	atgaagtata	atagtgcaca	24900
agtatagaac	ttgaaaatct	tccttaacct	taccataagg	gaaatgatta	ctaataagtt	24960
tcttaacttt	ttgtacttac	ataaacataa	atattcatca	gagaaaaaaa	tatgcaaac	25020
aatttgcaat	cttttttact	taccatattt	tggaattttt	ttcatttcaa	tatatttgat	25080
cttccttggt	tttttttcagt	ttgtttgttt	ttgtcaccca	ggttggagtg	aagtggtaag	25140
aacatggctt	attgcagcct	caatctcctg	ctcccattca	gcccctcaag	tagctgggac	25200
tacaggta	tgtcaccacg	cccggcta	ttttattttt	attttggtag	agatggggtt	25260
tcaccatggt	gcccaggctg	atcttgaatt	cctgggctca	agtgaccgcg	ccaccctcagc	25320
ctcccgaagt	gttgagagaa	cagggtgtga	ccaccatgct	ccacctctta	gtctttacaa	25380
tctgcaaaac	ctcataagtg	gctaatagag	gaatatagta	aagcaaaggg	ggatatcact	25440
gattagaact	gtgttttttag	gctgggtg	gtggctcacg	cctgtaatcg	caacattttg	25500
ggaggctaag	tgggagtatc	acttgagccc	gggagttaa	gaccagcctg	ggcaatatag	25560
tgagagaccc	tgtctttaga	aaaaaattaa	ccagggtgtg	tgggtgcacac	ctgtggtccc	25620
agctattcaa	gaggctgagg	tgggagaatc	gcttaagcct	aggaggcgga	ggttgagtg	25680
agatcatacc	actgctctct	agcatgggtg	acagagcgag	acccagtcta	aaaaaaaaag	25740
tattttttcg	ttttttttcca	actcatgtac	acccgccacc	ccaccctctg	tttttttttt	25800
tctgacattg	ggtcttgctc	tgtcaccacg	gctagagtgc	agtagcaaa	tcaactcact	25860
gcagcctcct	cttcctggac	tcaagcagtc	ctaccacctc	agcctcccaa	gtagctggga	25920
ccacaggtgt	gcaccatcat	gcctggctaa	tttttgtact	ttttgtagag	atagggtttc	25980
accatgtggt	tcaccatctc	taactcctgg	gctcagtcag	tccacttttg	cctcggcatg	26040
agccactgtg	tgacagccac	gtttttttatt	aatggatatt	tggattgttt	ccatctattg	26100
tgaataatgt	ggctatgaac	attgggtctaa	atatctgttt	aagtcccggc	tttcaatact	26160
tttgatata	tacctaggag	tagaattact	gaattatatt	gtaactttct	gtttaaattt	26220
ttgaactgcc	aacctgtttt	ccataggggc	tgcaccattt	tgcattccca	ccagcagtg	26280
acaagggttc	cagtttctcc	acatttggtta	tttttcattt	tttaaaataa	tagtcacct	26340
aaagggtatg	aagtgggtatc	tcgttggtgat	tttgatttgc	atgtattttt	ctaatagact	26400
atgatgtcta	gcattttgtc	atgtacttat	gtaccatttg	tgtatcttct	ttggaaaaat	26460
gtctattaat	gttcttttcc	catttttttaa	ttgggtgtgt	tttatgttta	tcaattttgt	26520
aaacatttta	agctctgcag	cataactact	caaccctgtc	acatggtaag	attgaccag	26580
taaaacttta	tgtacaaaaa	taggcagctt	actagattta	atcttagtcc	atagtttgct	26640
aaggcatgca	ttagataatg	tagttacact	attggcta	aatttaaa	acaagtgggt	26700
gtaagtttct	gccacccaaa	ttctttcttg	atttgatgta	gtctggttg	ttgaatttga	26760
gtgttatatg	gggtcataga	gttaaaagag	aaatgtctat	gagaaaactag	ggactgttg	26820
gagctaattg	taaaggattt	tggaggcctt	ttgtgcactg	gagaccattg	gaagattgga	26880
ttttctgcta	taaatgtatc	taaaggataa	tcagtgtaa	ttatgggctg	tagtttgcca	26940
accctacat	taagggttg	aaataatttg	aatatgggtt	tcagttcttg	tatgctctgg	27000
cctcagttct	tttttttttt	tttttttttt	tttttttttg	agatggagtc	tcactctgtt	27060
gcccagctg	cagtgcctatg	gcacgatctc	agctcgtctg	aacctccgcc	tcctggcttc	27120
aagtgattct	cctgcctcag	cctccagagt	agctaggact	acaggcacgt	gccaccacat	27180
ctggctaatt	tttttttttt	tcaagacaga	gttctgctct	tgtaaccag	cctggagtgc	27240
aatggcacga	tcttgggtca	ccgcagcctc	cgcctcctgg	gttcaagtga	ttctcctgcc	27300
tcagcctccc	aaatagctga	ggcaacaggc	gtgcgccacc	acgcctagct	aattttttctg	27360
tttttagtag	agatgggggt	tcaccatctt	ggccaggctg	gtctcgaact	cctgatctca	27420
tgatccaccc	gcctcagcct	cccaaagtgc	tgggattaca	ggcgtgagcc	acctcacctg	27480
gcgttttttt	gtgtgtgttt	tcagttagaga	tggggcttca	ctgtattagc	cagaatggtc	27540
ttgatctcct	gacctcgtga	ttcacctgcc	tcagcctccc	aaagtgcctg	aattataggc	27600
atgagccacc	acgcctggct	gtggctcagt	tcttaactgt	tcattcattc	agggtcccag	27660
ccaaaatctt	ttgccacagc	ttctcctggg	catgctctga	aattcatttt	ttgtctgttt	27720
agttgcatga	gtccactgaa	agctttgctt	attttctcca	ccactttcag	agttcacagt	27780
cacttcaagt	agacgaggtc	aggagatcga	gaccatcctg	gccaacatgg	tgaaccccg	27840
tctctactaa	aaatacaaaa	cttagctggg	cgtggtagcg	tgtgcctgta	gtcccagcta	27900
ctcaggaggc	tgaggcagga	aaattgcttg	aacccgagag	gcagagggtg	cagtaagccg	27960
agatcatacc	atttcactcc	agcctggg	acagagttag	acactgtcaa	aaaaaaaaaa	28020
aaaaaaaaaa	aggaaagaaa	agtgtctcaa	atgctgggct	catctctgct	ggcttctctc	28080



ctttttatat	ttcgataatt	tttaaaataa	taaccatcct	aatgggcttg	atatgggtccc	28140
tccttttctaa	atttttggtaa	ttttcattgt	tttggaact	cactaatccc	ttcaaacaga	28200
tgtttttaaa	actttttttc	tagttctctc	tctttttttt	tttttttttt	tttttgagac	28260
agggctctctc	tcagtggtgc	gacctccctt	actgcaacct	ctccctcctg	ggttcaagca	28320
gcgattctcc	tgctcagcc	tcccaaatag	ctgggagtag	aggcaccac	ccccagccc	28380
agctaattgtt	tgatatttta	gtagagattg	ggtttcatca	tggtggccag	gctgggtctcg	28440
aactcctgac	ccaggtgat	ccgcctgcct	cagccttcca	aagtgcctgg	attacagcca	28500
tgagccaccg	cacctggctc	tagttctctt	ctaactctac	tatagtaatc	aaaatttttag	28560
tgctgatgtt	ataatccaag	cccaaattgga	tattcttata	cattaaatgt	tggaatatca	28620
tgtttggttaa	aaatcagatc	tgctgggcac	ggtaggtcaa	tgctgttat	cccagcactt	28680
tgaggaggccg	aggcaggtgg	atccccctgag	atcaggaggt	cgagaccagc	ctgaccaaca	28740
tgagaaaact	ccgtctctac	taaaaataca	aaatttagccg	ggcgtgggtg	cacgtgcctt	28800
tcacccagc	tacttgggag	gctgaggcag	aagaattgct	tgaacctgtg	aggcggagggt	28860
tccagggagc	tgagatcgtg	ccattgtact	ccagcctggg	caacaagagc	gaaattccat	28920
ctcaaaaaaa	aaaaatcaga	tctgtcccta	tggttttata	ttcactgcat	gtcatatata	28980
aatggaagca	tgtagtatgt	agccttttgt	gtctcgcttc	tctccctttg	cataatattt	29040
ttgagattta	tcccttttcc	tacatatgtc	agtagtttgt	ttctttttta	tgctgaatag	29100
tattccacgt	gtggagatac	cacaatttgt	tactccattc	actaaatgat	ttgggtgtgt	29160
tcctgttttt	gttattgttg	ttgttggtgt	ttaattaatt	aattaatttt	ttttttgaga	29220
tgaggtcttg	ctctgttgcc	caggctagag	tgtagtggtg	caatcacagc	tcactgcaac	29280
ctctgcctcc	ccggttcaag	tgattctcct	gcctcagcct	cccaagtagc	tgtgattaca	29340
ggtgccgcgc	accacaacca	gctaattttt	gtatttttag	tagagatggg	atttcaccat	29400
ggtggccagg	ctggtcttga	actcttccct	ccttggctctc	ccaaaatgct	gggactacag	29460
gtgtgagcca	ccatgcccat	cctcctgttt	ttaaatttta	tgaataaagg	tactgtaaat	29520
tcagtacag	gtcttagaag	tttgcatttt	tcttgggtaa	atatgtagaa	gaagagattg	29580
ttgactcatg	tggtaaatat	atgtttaatg	tcataagaaa	ctaccaaact	gtttttccaa	29640
ctgggtgccca	ttttgtttcc	taccagcaaa	cataatgaaa	ggtccattac	cctttgtctc	29700
gtaatatttt	ggtattttta	tggtcttttg	tttttttgcc	attcaaaatt	gggtgcattg	29760
ttgtatcttt	ggattttaat	ttgtacttcc	ctaacaacta	atgatgtgga	gcactctttt	29820
atacgcttac	ttgccaatta	tgtcttaatt	ccttggcaat	acttatattt	gactctgcct	29880
actatatata	tatatatata	tatttttatg	ctatatatag	agratatata	tttttatact	29940
atatatatatt	atatatacta	tatatattta	tatatactat	atatatatat	atatatatat	30000
atattttttt	ttttttttga	gatgaagtca	tgctctgttg	accaggctgg	atgcagtggc	30060
atgatcttgg	cttaccgcaa	cctcaacctc	tgctcccag	gttcaagtga	ttctcctgcc	30120
tcagcttctc	gagtagctgg	gactacagac	acatgccacc	atgcccggt	aatttttgta	30180
tttttagtag	aggtggagtt	tcgccatgtt	gtccaggctg	gtcttcaact	cctgacctca	30240
ggtgatccac	ccgcttcagc	ctcccaaaag	tgctgggatt	acaggcatga	cccaccgtgc	30300
ccagctggaa	taagcaatct	taaaaagcag	tttgttgatt	tctggtgaat	agagaaaatg	30360
tgacaaacta	tgactcaata	gtttcatttc	ttttttgttt	gtttgtttgt	ttttttgaga	30420
cagagtcttg	ctctgtcacc	caggctggcg	tgtagtggtc	cactgcaage	tctgtctcct	30480
gggttcacgc	cattctcctg	cctcagcctc	cggagttagt	gggactacag	gcgcccgcga	30540
ccatgcctgg	ctaatttggt	ttttgtattt	ttagtagaga	cagggtttca	ctgtgttagc	30600
caggatggtc	tcgatctcct	gacctcgtga	tccaccctgc	tcggcctccc	aaagtgcctg	30660
gattacaggc	atgagccacc	gcccctggcc	tacttttatt	tcttaagcat	atatcttaaa	30720
gaaactgcag	gtatgcatta	ggggacaggt	agaaaaatgt	tacagcattg	tttgaaatgt	30780
tgaaaaactg	gaaacaacct	caatatctta	ataatagaaa	aatgggttga	ggctggctgt	30840
ggtggttcat	gcctgtaatc	ctagcacttt	tgagggtgta	ggcgggcgga	tcacctgagg	30900
tcgggagttt	gagtcagcc	tggccaacat	ggtgaaaccc	cgtctctact	aaaaatacaa	30960
aaaatttaggc	aggtgtgggtg	gtgcacacct	gtaatcccag	ctacttgga	ggctgaggca	31020
caagaatcac	ttgagcctgg	gaaatggagg	ttgtagttag	ccjagatcat	gcctactgca	31080
ctccagcctg	ggcaatagag	tgagactccg	tctcaaaaaa	aaaaaaaaaa	aaaaaaaaagg	31140
aaggagcatg	ggaatgggaa	gaggcactaa	aaaagggaact	tcttttgat	ctgttacatt	31200
gtactcattt	tgtttttaaa	attttaagcc	aaatatgaca	ttgttatcac	ttttgtgcc	31260
actcagaata	caagtgtctat	tacactattt	tctataattt	tccctttttc	aaactaaaaa	31320
acaaaaatac	atgtaaacag	gtgttttctt	tatggaaaac	aggaagggtga	gcacccaaat	31380
aatttttagtt	aattttataac	aaactcttaa	taattttctt	ttcttcttct	tctttttttt	31440
tttttttttt	tgagtccagg	tcttgccttt	ttgtcagggt	tagagtgcac	tggtacaatc	31500
tggtacaatc	acatcttagt	gcaactctta	cctcttcagc	tcaagtgacc	tctctcctca	31560
gtcccccacg	tagatgggac	cacagatgtg	tatccccaat	catggttaat	ttttaaattt	31620



tttggttagag	ataaaatctc	accatgttgc	ccaagctggg	cttgagctcc	tgggctcaag	31680
cagtcctcct	gccttggcct	cccaaagtgc	tgggattata	ggcgtgagca	aactgtgccc	31740
agtcttcata	aacttttctt	tacatgtcct	tatcaagtac	tttttgagca	cctactgtca	31800
accagccgta	gtatataata	ctgataattc	tataacataa	gaaattgacc	tgtttaaggg	31860
gatgaatatg	gttagttatt	ttcagggtga	agaaacagaa	gaatcgggga	ggtagtacat	31920
agtcataagg	agacggcatt	tcttagtcac	tttgtgtggt	gtttataata	agacttcatg	31980
tggccagggtg	cggtggctcg	tgagcgctcg	taatcccagc	actttgggag	gccgagggtg	32040
gaggatcaca	acaaggtcag	gagatcgaga	ctatcctgaa	caacatgggtg	aaaccccgctc	32100
tctactaaaa	atacaaaaat	tagctgggtg	tgggtggcacg	tgcctgtaac	cccagctact	32160
cgggagggtg	aggcaggaga	atcgcttgaa	ccagggagtc	ggagggttgca	gtgaaccaag	32220
atcggtgccac	tgcactgcag	cctgggctac	agagagagac	tccatctaaa	taaataaata	32280
aataaataaaa	taaatagact	tcgtctgttt	catcaagagt	cattgtatta	tattgatttt	32340
gaaatttcaa	ataataataa	ggaatagtat	ggaaattttt	gtttgattaa	atggggacat	32400
gaggccagggt	gcgttcactc	aggcttgtaa	tcccagcact	tcgggagacc	aggcgaggag	32460
gattacttga	gcctaggagt	tcaagaccag	cctgggcaac	atggcaagac	cctatctcta	32520
caaaagaaat	acacaaaaat	tagctaagtg	tgggtggcaca	tgcctgtagt	cttagctact	32580
cgggaggcca	agggaagagg	atggatcact	tgagcccagg	agttggaggg	tgcagtgagc	32640
taggattgtg	cccctgcact	ctagtgtggg	tgacagcaag	actctgttta	aaaaaaaaaa	32700
aaaatgaggc	cgggcgcggt	ggctcacctc	tgtaatctga	gcactttgag	aggccaaggc	32760
gggtggatca	cctgagttca	agatcagcct	ggccaacatg	gtgaaaaccc	cgtctttata	32820
cttaaatatc	aaagattagc	tgggtgtgat	ggtgcacacc	tgtaacccca	gctactccag	32880
aggctgagac	agcagaatcg	cttgaacccg	ggagggggag	gttgcaatga	gccgagattg	32940
caccactgca	ctccagccta	ggtgacagag	caagactctg	tctcaaaaaa	aaaagaaaaa	33000
aagtctaaga	agtttaatttt	cattcagaca	aatgttcaaa	aataatagaa	ataaaacaaa	33060
aaattaattt	tcaaataatgt	gatttgagta	taacgatcac	tttacagaag	ttattctaca	33120
tactttttga	ttatgctgag	tttttaggtc	cttttaagc	ttaacttgta	tcaggaatta	33180
atgtttttaa	ctagtgtgtg	ggttcaaaga	aaatagaaat	gtgttattaa	atgccacact	33240
tgtaatctca	tacagcagtt	ctcatgatgc	aatcaattgt	taaataaaaa	cttcctctag	33300
agaggaggta	cccagcaatg	aacacattat	atcagatttg	aatatgagat	taaacaatgc	33360
tttttgtttc	attgttattg	aaatcatatt	tgtatttttc	aaaagtatat	atacttaaaa	33420
attcctaata	aaaatgtctt	tctgtaatac	agcatttttg	gttataagga	tcaataccat	33480
gggttgataa	gcaaaatagg	actcacggcc	ccaaaatggt	gataaacatc	atgaccatat	33540
tcccataaat	ggagttacat	atacacatta	caatgataaa	aatactaaga	agttatatag	33600
taatgtttct	cagacttggt	cacaaccctg	ttttttatgt	attacctctc	aaaagcatga	33660
tttgcaatat	ttagtgtact	cttcatacga	tctatacaaa	taatttttta	ttttaagca	33720
ggacagaagg	agcaagctgt	ggaatggtat	aagaaaggta	ttgaagaact	ggaaaaagga	33780
atagctgtta	tagttacagg	acaaggtaag	attgtatttg	tttatagcca	tcccaaatta	33840
tgatatattc	acatgattgt	ccagatttca	gatctattta	tttatttatt	tatttttctt	33900
tctttctttt	cttttctttt	cttttttttt	tttttttttg	agacagggtc	tttctcttgt	33960
taccaggtct	ggagtgcagt	gatgtgatca	tagctcactg	caacctcaaa	cttctgggct	34020
caagtgatcc	tcctgcctct	gcttcccatg	tagctgggac	cacaggcgca	cactaccata	34080
cttagctaat	tattattatt	attattatta	ttattattac	tagtttttga	gacagagtct	34140
cgctcttttg	cccaggctgt	agtgcaatgg	tgcaatctcg	gctcactgca	atctccgctt	34200
cccggattca	ggctatttat	tctcctgcct	gagactcctg	agtagctggg	attacagggtg	34260
catgccaccg	tgcccagcta	attttttgtat	tttttagtgaa	gatgggggtt	tgccacgttg	34320
gccaggctgg	tttogaactc	ctgatctcaa	gtgatctgcc	cacctcagcc	tcccaaagtg	34380
gtgggagtag	aggcatgagc	cactgtgggc	cacacttagc	taattaaaaa	aaattttttt	34440
gtagagacag	ggtctcactg	tgtttctgtg	actgatctta	aacttctggc	ctcaagtgat	34500
cctcctgcct	cagctttcca	aagttgtggg	attacaggca	tgagccactg	taccagcca	34560
gattgtattt	atcttaataa	atgtatttgg	gccaggcaca	gtggctcatg	cctgtaatcc	34620
cagcactttg	ggaggccgag	gtgggcgaat	caactgggat	caggagtctg	agaccagcct	34680
gatcaacatg	gagaaacctt	gtctctacta	aaaattcaaa	attagccagg	cgtgggtgtcg	34740
cgcacctgta	attgcagcta	cttgggaggc	tgaggcagga	gaattgcttg	gaccagggag	34800
gaggagggtg	tggttaagcca	agatcgcgcc	attgcactcc	agcctgggca	acaagagtga	34860
aactctgtct	cagaaagaaa	gaaaagcaaa	acagtatttc	atttaattaa	aaatgccttg	34920
tattatacca	gtaagagtaa	aaaacaattt	aattaaagta	tgatatatgc	tcttttcatc	34980
acttagaatt	tgaatttttaa	tcttattgaa	gaggctgttt	tagttttatt	tagatatggt	35040
tataacaaca	atattaatat	ctgtgaccaa	attggtgcat	gcatgggcaa	tgaaaagcta	35100
tgttgtgact	gcagcatggt	ggatagcatt	agtaagttag	cattggttgt	aaaatgaatc	35160



ccaatttcaa	agattttgaa	acgtggggaa	aaaatatattt	agaccccatg	aaataagacc	35220
tgaaatctat	gaaataccac	agtaggttgg	aatcatcat	gagaaactgt	aactattttt	35280
ttttataggt	attatatgta	ttctttgaga	cagggctctca	ccctgttgcc	caggtctggaa	35340
cgctgtgatg	tggtcacctc	tcactgcaac	atccacttcc	tggggctgaa	gtgatccttg	35400
caactcacc	tcccatgctc	ggctaatttt	ttttttttt	tttttaagta	gagaccagat	35460
ctctttatgt	tgcccaggct	ggtcttgaac	tcctgggctc	aagtgattct	gctttgtctg	35520
cccagagtgc	tgggattaca	gaagtgaacc	acaacacctg	gcctaaagat	agtattttta	35580
aaattttctt	tgttttcttt	cttttggtta	tacattttct	tctttttttt	aaaaaaaaat	35640
ctctgtataa	agactgtgac	tcccatgaa	agtagtttgg	gtgataattt	atcgtgaaac	35700
tattagtt	gggaaatgta	gatattttta	ttaatttttt	tctttcaggt	gaacagtgtg	35760
agctag	acgccttcaa	gctaaaatga	tgactaattt	ggttatggcc	aaggaccgct	35820
cttct	aggtatcaat	taatgtataa	tttgatgtgg	gatgtattgg	aaatgtgtgt	35880
aaac	tttaatttgt	agaaagaaat	agatcagtga	ttgaaaatgt	ggtccaggct	35940
ta	aagaaaacgt	ataacatata	caaaaacaga	agagcataat	ggacttcttt	36000
	ctgggacgct	tgctctgtca	cccaggctgg	agtgcagtg	cgtgatcttg	36060
	ctacct	cccaggttca	cgccattctc	ctgcctcagc	ctcctgagta	36120
	cccccg	ccaccacgcc	ccgctaattt	ttttgtagtt	ttagtagaga	36180
	ctagc	cagaatgggtg	tccatctcct	gacctcgtga	tccacccgcc	36240
	agtgctgg	agttacaggc	gtgagccaac	gcgcccagcc	gagcataatg	36300
	tattcttctc	tcagctttga	taattatcac	ctcactatca	gtcttgattt	36360
acca	agggttagca	aactgcagct	gatgggcca	cccagctcac	tgccatttac	36420
ataag	gttttattgg	cacacagcca	ttcctttcct	ttaggtattg	tctatggctt	36480
agctacg	gtgacagaat	tgaatagttg	atacagagac	aacactgccc	tcaaaaccta	36540
catatttac	tgtctggccc	tgagaaagcc	tgccatcttc	acttgacatc	ccacttcaat	36600
gtcccccaac	ttcttattat	tttgaacaaa	attattgact	tattttccat	ttataaatat	36660
ttcatattat	acctctaaaa	gataaggatg	tcaaaataaa	atcacagaaa	taccattatc	36720
acacctaaaa	aataactcca	aatttctttt	aacattatcc	aaaaaaactg	acccccccaa	36780
aatgcaattt	ccaaatttct	aattttctca	aaatttaata	atttttacta	ttttttaata	36840
atttgtttgt	atcaggatct	taaaaagatc	cgtatgtctc	ttttaatgta	caggttcac	36900
tgtttttcat	tttcttccaa	tttatttgtt	gatgaaacca	ggtctttgtc	atgtaatat	36960
tcctacagtc	tgacttttgc	tgcatgcac	ttttgtggta	tgttttaaca	tgttcttctt	37020
tctctccacg	ttaaatgata	gttgatata	gagccttgat	cacattgaaa	gttgattttt	37080
attttttaat	gattagacta	cttctagggt	ggtgtgtgtt	gtgtgttcat	cattagttgt	37140
tccatcttgt	gaggttagca	gcagtttaact	atgactgatg	tccagtata	ctttaattct	37200
gtcattcttt	cagcattcat	caactgtaat	agaaactttt	acttgtcttc	tgtttgatgg	37260
cgtaggtttt	aaatatatat	taataagtta	aatctaccac	tcaaaaaaag	tagaacaata	37320
cctactacat	atgagtggct	ctaccatacg	ttgggtggatg	gaaatagtag	atatttggtta	37380
agcaggtaaa	ccttatttat	ttgtagccct	gtgacttggg	gtaagtattt	ttgcatgtct	37440
gggtcttttt	tatctaattt	gtaatatgaa	gctattatta	ttattattat	tattattatt	37500
attattatta	ttattattat	tattcgagat	ggagtcttgc	tttgttgccc	aggctggagt	37560
gcagtgggtg	aatctcggct	caatgcaccc	tccacctcct	gggttcaaac	tattctcgtg	37620
attcagtcct	ccgagtagct	aggattacag	gggcctacca	ccacacctaa	tttttgtatt	37680
tttagtaggg	acggggtttc	accatggttg	ccaggctggt	ctcaaaactcc	tgacatcaag	37740
tgatccatcc	acctcagcct	cccaaagtgc	tgggtagggc	cgggcacggt	ggctaacgcc	37800
tgtaatccca	gcactttggg	aggccgaggc	gggtggatca	cgaggtcagg	agattgagac	37860
catcctggct	aacacggtga	aaccccgctc	ctagtaaaaa	aatacaaaaa	actagccagg	37920
tgtgtgtggc	ggcacctgta	gtcccagcta	cttgggaggc	tgaggcagga	gaatggcatg	37980
aaaccaggag	gcggagcttg	cagtgaagcc	agatcatgcc	actgcactcc	aggctgggca	38040
aaagagcgag	actctgtctc	aaaaaaataa	aataaaataa	atagataaat	aataataata	38100
aagtgtctgg	attacaggca	tgagccacca	cgcccagctg	aagctaatat	tattagctaa	38160
taataatata	tcattatttt	gaccattgaa	ttatttaaagg	aacctaaacc	atagtagtaa	38220
gtgctttaaa	aatgttgggt	tttattttat	tgtttacctg	actgtcatca	cctcctttgc	38280
ttaggtagaa	agatatattc	ataaacacat	acatatatcat	tttattcatt	ctttatttca	38340
tttactccta	tatggtcatt	gtgtttgtgg	gattttttgcc	cgcacatagt	taaaaatgca	38400
gccgttactg	ctttcagatg	agtaacaagg	tagtgttccc	tggcttctgt	ggctgacagg	38460
atttgccttc	ttcctttcta	aatggaggtt	attactgtgt	cagatataat	taaatagtgt	38520
attaaagttt	gttgcaataa	aataattgat	ggttctaatt	ggtactttct	acgtgtttta	38580
tctttaaagc	ttttcagttg	atatataaag	tatatatcat	acagaataaa	tttgattgtg	38640
gagcattttg	taacatcttt	caaaattaat	aagagttggt	ttttatgttt	tgtttgatgt	38700



.

.

.

.

tatgatatag	gacattagaa	gtattaatat	caattagaga	ctcatctttg	aatgtgactt	38760
tgtactttct	tattttgtgt	agtagaggag	agaacaaaaa	gaagatatgt	aatgtaatat	38820
gaaactaggc	atttaaagat	ttaacgtttt	ggatatttta	aagttgggtg	ctgttttcac	38880
cctcaaaaat	gctattacct	atztatgaaa	tatcttttaa	agtgtggagt	ggtagatggg	38940
aaaggtgaca	tcggctgggc	gcggtggctc	acacctataa	tcccagcact	ttgggaggcc	39000
aaagtgggcg	gatcacctga	ggtcaggagt	tcacgaccag	tctggtaaca	tggtgaaacc	39060
taatctctac	taaaaataca	aaaaattagc	caggctgggtg	gtacgcacct	gtaatccag	39120
ctactcagga	ggctgagtc	ggaggatggc	ttgaacccag	aaaatggagg	ttgcattgag	39180
ccaagatggc	accactgcac	tccagactgg	gtgacagaga	gagactccgt	ctcaaaaaaa	39240
aaaaaacgcc	gggcactgtg	gctcacacct	gtaatcctag	cactttggga	gactgaggcg	39300
agcagatcac	ttgaggtcag	gagttcgaga	ccagctgggt	gcctgggcaa	catagtga	39360
ccccgtctct	actaaaaata	caaaaaaatt	agccggggcg	agtggcgggc	gcctgtagtc	39420
ccagctactt	gggaggctga	ggaaggagaa	tggcgtgaac	ccgggaggcg	gagcttgcag	39480
tgagccgaga	tcccgcact	gcactccagc	ctgggtgaca	gagcgagact	ccgtctcaaa	39540
aaaaaaaaaa	aaaaaaaaaa	tagccgggca	tgggtggcgg	cacctcta	ctcacctact	39600
tgaggaggtt	aggcatgaga	attgctcgag	cctgggaggc	agaggttgca	gtaagcagag	39660
accacaccac	tgcactctag	cctggatgac	agagtggagc	tttgtctcaa	aacaaaaaca	39720
aacaaaccaa	accacctctt	gtccactcct	aatctccctt	actgtgtttc	acttattttt	39780
ttcccatata	tttgtcacct	tctaataaac	tgtgtaaact	attttttatg	tttattgtgt	39840
actatctgtc	ttctccctgt	tggaaataaa	actctatgga	ggagggatcc	ttgtctatat	39900
cccaagtggc	tacaacagt	tctagcacat	gctagttacc	gaataaatat	ttgccaaagt	39960
aatgaagtgt	tttataaccc	ttggcactat	acctgataca	gtgtagggac	ttaataaatg	40020
aatgttaaat	gaagctatct	taaaaatata	tatatatata	taacagccag	atacgccaat	40080
ttgcccagcat	caaaattaat	cttttcatct	attgcaggtt	ttacagttaa	gctttgtctt	40140
aaggttcaact	atcactaaac	caaaaataaa	taatatggat	ggtcttttcc	catttttaac	40200
agagaggcta	catcctctca	aaccagtcag	tattaggaat	tcccaggatc	attctctcca	40260
ttttccatcc	taattagata	cctatctggc	accacatctc	tgaaaagctc	caatttttaa	40320
ccagcacata	ataaacaat	atgttactat	taatacacgt	ttgtacacaa	gtcattttta	40380
ttgtcttagg	tatatccact	cccagacctt	ctgagttatt	gaaagacagg	aattttgtgat	40440
gtcacctctt	ttttcttttt	tttgagacgg	agtcttgcc	tatcacccag	gccagagtgc	40500
aatggcgtga	tctcggtcca	ctgcaacctc	tgcctccgg	gttcaagcga	ttatcctgcc	40560
tcactctcct	gagtagctgg	gattacaggt	ggccttagtt	tgggtttttg	gttttaaaaca	40620
ggatggtctc	ctctttgaga	agttgccatg	agttagccat	gatgatacct	gggacaataa	40680
tgggtccacag	agaagtaaga	gccaatacaa	cgccttaaga	gaaagtacct	ggtatgtttg	40740
tgaagcagca	aagaggccaa	tgtgttccag	caaggagatt	atagtagaga	tgaagtcaga	40800
gaactcatgc	ggagagggat	tgtcacacaa	cgcagggtcg	taaaggccac	catgaggact	40860
tgagctttta	ccgcctgtga	aatggggagc	agttacaaag	atttttaaga	aagcagccca	40920
gtgcagtggc	tcatgcctat	aatcccagct	ctttgggagg	ctgaggtggg	tgatcacctt	40980
gaggtcagga	gttcgagacc	agcctggcca	acatggcgaa	accttgtctt	tactaaaaat	41040
agaaaaatta	gctgcgtgtg	gtggccatgc	ttgtaatccc	aaatactcgg	gaggtgagg	41100
cagaagaatc	acttgaaccc	gggaggcgga	ggttgacgtg	agccgaggtt	gcgccactgc	41160
cctccagcct	gggcaacaaa	gcgagactcc	gtctcaaagc	cagaaaaaaa	gatttttagg	41220
aaaggaatga	aatactctga	cttaaatttt	agaaggatca	ctctggttgc	tatgttgcat	41280
tatagactct	aagggtagaa	ccagggtag	caattaggac	actgttacag	taatctagga	41340
gaaaagtgat	agaggcttgc	accagggtag	tatcaggaga	aatggtaaca	agtggttgga	41400
ttatgattat	agttgaaggt	cgaacaacta	gatttgctga	tagattggat	aaaaatgggt	41460
tctctttatg	gatctcataa	tacagttaga	gaaacagaca	cataaatagt	catactgca	41520
ggatgggtat	tagaaatatt	ttgcaagggt	gtaaaagggt	ctaagaaagc	tttcttgtgt	41580
gacaaagctg	gttaaagaaa	gaaaagaaag	cttcctagaa	gtgaaaccta	aaatatacac	41640
ttagataaat	gaataaaagt	tatgaaagaa	tgtgagggct	gggcacagt	gtcctacact	41700
gtaatcctag	cactttggga	ggttgagtcg	gggttattgc	ttgagtcag	gagtttgaaa	41760
ccagcctgga	cagcgtactg	agaccccatg	tctaaattta	aaaaaaaaaa	aaatagccag	41820
gagtggtgat	gtgcacctgt	agtcctagct	acttacttgg	gaggctgagg	caggaggatt	41880
gcttgagctc	aggaggtcga	gactgcagt	agccatgagt	acaccactgc	actctagtct	41940
gggtgacaga	gtgagaccct	gtctcaaaga	aaatgaaaag	aaaaagaatt	tgagaaaaga	42000
tatgtcagga	aaatgtgatg	acatgaacaa	atactcaaa	gcaagaaaag	catggtgagt	42060
aggggagggg	acaggttgag	acatgaagtt	gggattgccg	agggcgtaaga	ggatataggca	42120
gatcatggat	ggcctggttt	gtcataggaa	ggagcttggg	ctcttatctg	tgggcaatgg	42180
gaagccacta	aagggtttta	agtagaagag	tggtatatgg	taagggttttc	ctctccagta	42240



gatcactcag	atgactgtgg	gggttggatg	tgaggaggta	aagcagcaaa	gttactgctt	42300
taattttaa	cattgattaa	tctgtacaat	cctagtcatt	ccaaaaagaa	acattagtcc	42360
tttaacagaa	gtgaattggg	tctgggcgcg	gtggctcacg	ccgtgaatcc	cagtactttg	42420
ggaggccgag	gcgggcagat	cacctgaggt	caggagtttg	agacaagtct	ggccaacatg	42480
gcaaaacctt	gtatctacta	aaattacaaa	aattagctgg	gcatgggggc	acccacctgt	42540
aatcccagct	attcggggagg	ctgaggcagg	agactctctt	gaaccataaa	ggcagagggt	42600
gcaatgagct	gaggtcgcac	cactgcgctc	cagcctgggc	gacagagtga	gactccgtct	42660
ccaaaaaaaa	gaaaaaatat	tggagcagtt	tcacagatgc	tgtttactgt	tatactgtat	42720
gtgtctatga	ctcctcctcc	aagaaaaaaa	aaatgaattg	gagcaggttc	acagaagcaa	42780
tgtactgtta	cactgtatgt	gtctgtttct	acacataata	atctgaattc	tgtgtacacg	42840
aaaagaaatc	ctatagcttt	ttattcctag	ctataaaaac	taagaatata	atttctttct	42900
ttttattaca	taatgaatgg	ttctgttaac	tttttgttaa	aatttcattg	aggggagtat	42960
taattcacat	ttatgtacgg	atgtctacat	ttacaaatca	gtgtattttt	tgatttatgc	43020
ttttactgag	acaaagggtt	tctgtctcag	catggtcatt	taaagagttt	atcattgaga	43080
aaaatcagat	gaccaacctg	ttagctcaaa	aaaaaaaaaa	cctccaaggt	atatgtatc	43140
agccagttct	aggatacaaa	agccatgcag	tactttgtgt	tttgtgccaa	aaagggttagc	43200
tgtctattga	cctgtcccaa	aggcatgtgt	ggttgtaccg	taaaccaagc	atggtacctg	43260
tttgtcaaac	tttagaaatg	aaagttaag	agagttaata	tataggtgct	gcatttttta	43320
tgtattcatt	gacttgctgg	tacagaagaa	aagaatcaat	tatgattcag	cacaatactc	43380
cacttgggga	agagagtgcg	gcagtagttt	agagtgtcag	ggatcaaact	gctaccttct	43440
tgggcttcag	ttgctagact	taagagaccc	agatcttggg	aggggttttt	gttggtattt	43500
gatgtggggg	taaggttgaa	atactccctt	ttaattgatt	aataataaaa	ataggtaatg	43560
tagacgtatg	tcacaaaatt	ttaacaagtt	tgagtgtgaa	aagcatccat	cccattctat	43620
gccctggcta	cccatttccc	ctccctagag	gcggctgtta	ttatcagttt	ctctgggtcc	43680
tttttagagg	taatcttttt	gtatgcaact	gtgcacatgt	atgtttacac	aaatggtagg	43740
gtattgtata	tatatacagt	tactgcagta	tgcttatttg	actttgcttt	attgctttta	43800
ttttctgcct	gacatatggt	aatgtggcca	aatgggcat	ttcccagaca	atccttatag	43860
tttcattctg	tgctacttta	aatgctgttc	tccttccctg	aaatgtcctt	tttttttttc	43920
ttttcttttt	ttttttttga	gacagagtct	cgctctgtca	cccaggctgg	agtgcagtgg	43980
cacaatctca	gctcactgca	acctccacct	ccccggttta	agccattccc	ctgcctcagc	44040
ttcccagagta	gttgggacta	caggcgctg	ccaccacgcc	cggatacttt	tttgtttggt	44100
tgtttttagta	gagacagggg	ttcactgtgt	tagccaggat	cgctcaatc	tcctgacctc	44160
ctgatctgcc	tgccctcagcc	tcccaaagt	ctgggattac	agacgtgagc	cactgcgctt	44220
ggctgtcctt	tcttttaatc	ttttcaaatt	ctagccagtt	catagcccac	ttgctctgta	44280
aagctttttc	agtgatcttt	cttttgtacc	atgcagctct	ttaaaaaatc	tacagtttta	44340
ttataagttg	ctaattatat	ttaataaata	cctaaatggg	agaaattgat	tcatttttta	44400
atctccgtat	tgccaggaaac	tgtaggtcat	gtctgtaatc	cctgcacttt	gggaggctaa	44460
ggcgggagga	tcgcttaagg	ccaggagctt	gagcccagcc	tgtggaacat	attgagaccc	44520
cgtctctaca	aaaaataaat	tagctgggca	tagtggtgtc	tgtatgtagt	gtcaactact	44580
caggagactg	aggcagaagg	ataacttgag	ctcaggaggt	tgaggccaca	gtgagccatg	44640
atggtgccac	tgtacttcag	cctgggtgac	agagcaagac	caagaccgat	aattatctcc	44700
atggtatttg	ttttggacgt	attagatatt	cagtaaatat	tttctggtga	tgatagtgat	44760
atctaactac	ctcaggttac	ctaacctgaa	taaggcctat	ttttaacata	gccattcact	44820
tgtttaatttc	tgttattttc	gtgacttcat	ttttactaat	tatattttaa	aagataattt	44880
tatataaatg	caaaaacttt	ttatcatgta	acaatctggg	aacaccttga	gtaatttgte	44940
atttcacatg	cacattttat	ttgttcatta	tcttttttct	ttttgtttat	tttttctggt	45000
ttttaccttc	tctgttgcat	agagaagatg	caaccagttt	tgccattttc	caagtcacaa	45060
acggacgtct	ataatgacag	tactaacttg	gcatgccgca	atggacatct	ccagtcaggt	45120
gggttttaggt	taactaacat	aaaataataa	agcttgcatg	caaagtaaga	gtcttactta	45180
acctgataat	gttgatttga	ttttataatg	gtaggtttta	ttgttcatgt	tttcacaggg	45240
ctgtgttgaa	aatatagtac	ctttatcatc	atagtatata	aacatgcaac	aagtcaggta	45300
tagtcattct	aaacttgctt	tatttgcata	ttttgcaacc	tctggttttt	cagattataa	45360
aatatgaagg	tatacatctc	cgtaagtgat	taaaataatc	tttttttctg	ttgtggcttt	45420
tgtgttttta	tatggatagc	tatactgatt	tttttcatga	aatagtgttt	tctaagacac	45480
attttaattt	tttatgtatc	aattttttat	gcatatcaat	tttcatttat	agtatgtgtg	45540
aatgtcttat	agtacttgat	agaagaaatt	tgaaacttca	catagtgagg	agaaacatta	45600
cagtattatt	tgctatggat	taactctatc	cttcataaac	ttttggatat	caacttctac	45660
agaagattaa	atgtctttct	gttttgtttt	ctttgggtag	gggactctcc	atttcttagg	45720
ttctctgaca	tctgagctag	ctttttgttt	aattaaaact	ttttttttag	ttgacagata	45780



.

.

.

.

atgtacctgt	tcatggggta	cacagtgatg	ttttgataca	tatagtatat	ggtgattaga	45840
tcaaggtaat	tatcattttcc	atcatcccaa	acattttatca	ttccttgtgt	tggaaacatt	45900
cagtatttttc	cttctagcat	ttgaaatgat	ataatatatt	atattgttac	ctataatcat	45960
cctgcagtga	tgtagaacaa	tggagcttac	tctttctatc	tagctatacc	tttgatccct	46020
ttccctaccc	ttttcagcct	ctagtatcct	ctgttctact	ttttatttct	atgagattaa	46080
ctttttaaaa	ataatattct	gtgtgtatat	ataccacatt	ttctttattc	attcgtctgt	46140
tcttggacac	ctaggttgat	tctgtatcct	ggctgttggtg	aatagtgtctg	cagtaaacad	46200
ggggtgtaga	tgtctgttcg	atatgatgat	tttctttcct	ttggataaat	tcccaatagt	46260
gggattgggtg	gggtcatatgg	tagttctgtt	tgtagttttt	gaggaacctc	catactcttc	46320
tccatagtag	ctgtactagt	ttagattccc	aaaagtagcg	tataagagtt	cactttctac	46380
acatcctcac	cagcattttg	tacttttttt	gtctttttga	taatagccat	cctaactggg	46440
gtaagatacc	ttgttgtggt	tttgattttgc	atttccctca	taattaacga	tgttgagcat	46500
tttttcatat	ctttgttttt	gacggtccag	ctagttttat	tgtgtacttt	tttttttttt	46560
tttttagcaat	agtttcatgt	gcagggctca	aattatattt	aatattactt	ttttcaaatt	46620
ataccagaac	aaccttaggt	taggctataa	aacactgccc	ttctttttct	tttggcataa	46680
taattatata	tatatattta	attacacttt	acgttctagg	gtgcatgtgc	acaacatgca	46740
ggtttgtttac	atatgtatac	aagtgccatg	ttgggtgtgcc	gcacccatta	acttgtcatt	46800
tacattaggt	atatctccta	atgctacccc	tccctctctc	cctcacccca	cgacaggccc	46860
cattgtgtga	tgttccccac	cctgtgtcca	agtgttctca	ttgttcagtt	cccctctgtg	46920
agtgagaaca	tacggtgttt	ggttttctgt	ccttgcgata	gtttgtctag	aatgatgggt	46980
tccagcttca	tccatgtccc	tacaaaggac	atgaactcat	cctttttgat	ggctgcttag	47040
tattccatgg	tgtatatgtg	ccatattttc	ttagtccagt	ctatcattga	tggacatttg	47100
cattgattcc	aagtctttgc	tatttgtaat	agtgccacag	taaacataca	tgtgcatgtg	47160
tctttatagc	agcatgattt	ataatccttt	gggtatatac	ccagtaatgg	gatggctggg	47220
tcaaatggta	tttctagttc	tagatccttg	aggaatcgcc	acactgtctt	ccacaatggg	47280
tgaactagtt	tacgttccca	tcaacagtgt	aaaagtgttc	ctgtttctcc	acatcctctc	47340
cagcacctgt	tgtttcctga	gttttttaatg	atcgccattc	taactgggtg	gagacgggat	47400
ctcattgttg	ttttgatttg	catttctctg	atggccagtg	atgatgagca	ttttttcatg	47460
tgtctttttg	ctgcataaat	gtcttctttt	gagaagcatc	tgttcatata	ctttgcccac	47520
tttttgatgg	ggttgtttga	tttttttctt	gtaaattttt	ttaagtctct	tgtagattct	47580
ggatatcagc	cctttgtcag	atgggtagat	tgcaaaaatt	ttctcccat	ctataggttg	47640
cctgttcact	ctgatggtag	tttcttttgc	tgtgcagaaa	ctcttttagt	taattagatc	47700
ccatttgtca	attttggett	ttgttgccat	tgtcttttgg	gttttaggca	tgaagtcctt	47760
gcccatgcca	gtgcccagtc	cagtgtcctg	aatggatttg	cctagatttt	cttctagggt	47820
ttttatagtt	ttagaactaa	catttaagtc	tttaatccat	cttgaattaa	tttttgata	47880
aggtgtaagg	aagggatcca	gtttcagctt	tgtacgtttg	gctagccagt	tttcccagca	47940
ccatttatta	aataggaaat	cctttcccca	tttattgttt	ttgtcaggtt	tgtcaaagat	48000
cacatggttg	tagatgtgtg	gtattatttc	tgagggtctt	gttctgttcc	attgggtctat	48060
atctctgttt	tggatcagat	accatgctgt	tttgattact	gtacctctgt	agtatagttt	48120
gaagtcaggt	agcacgatgc	ctccagcttt	gttcttttgg	cttaggattg	tcttggcaat	48180
gcaggctctt	ttttggttcc	atatgaactt	taaagtagtt	ttttccaatt	ctgtgaagaa	48240
agtcatttgt	agcttgatgg	ggatggcatt	gaatctataa	attaccttgg	gcagtgtggc	48300
cattttcatg	atattgattc	ttcctatcca	taagcatgga	atgttcttcc	atttgtctgt	48360
gtcctctttt	atttcattga	gcagtggttt	gtagtctctc	ttgaagaggt	ccttcacatc	48420
ccttgtaagt	tggattccta	ggtattttat	tctctttgaa	gcaattgtga	atgggagttc	48480
actcatgatt	tggctgtttg	tctgttattg	gtgtatagga	atgcttgtga	tttttgaca	48540
ttgattttgt	atcctgagac	tttgcgaag	ttgcttatca	gcttaaggag	attttgggct	48600
gagatgatgg	ggttttctaa	atatacaatc	atgtcatctc	caaacaggga	caatttgact	48660
ttcttttctt	aattggatac	cctttatttc	tttctctctc	ctgattgccc	tggccagaac	48720
ttccaacact	atgttgaaca	ggagtgggtga	gagagggcat	ccctgtcttg	tgtagctttt	48780
caaaggggaat	gcttccagtt	tttgccatt	cagtatgata	ttggctgtgg	gtttgtcata	48840
aatagctctt	attattttga	gatacatccc	atcaatacct	agttttattga	gagtttttag	48900
catgaagggc	tgttgaattt	tgtcaaaggc	cctttctgca	tctattgaga	taatcatgtg	48960
gtttttgtct	ttggttctgt	ttatatgatg	gattacgttt	attgatttgc	atatgttgaa	49020
ccagccttgc	atcccaggga	tgaagccaac	ttgattgcgt	tggataagct	ttttgtgtg	49080
ctgctggatt	cggttttgcca	gtattttatt	gaggattttt	gcgttgatgt	tcatcaggga	49140
tattgtgcta	aaattctctt	tgttgtgtct	ctgccaggct	ttggatcag	gatgatgctg	49200
gcctcataaa	atgagttagg	gaggattccc	tctttttcta	ttgattggaa	tcatttcaga	49260
aggaatggta	ccagctcctc	tttgtacctc	tggtagaatt	cagctgtgaa	tccatctggt	49320



.

.

.

.

notggacttt	ttttgggttag	taggetatta	attattgcct	caatttcaga	gcctgttatt	49380
ngtctattca	gggattcagc	ttcttcctgg	tttagccttg	ggaggctgta	tgtgtccagg	49440
tatttatcca	ttcttcctag	attttctagt	ttatttgagt	agagggtgtt	atagtattct	49500
tgaaggtag	tttgcaattc	tgtggaatcg	gtggtagatat	cccccttata	atttcttatt	49560
catctattt	gattcttctc	tcttttcttc	tttattagtc	ttggtagcag	tctatcagtt	49620
tqragatct	tttcaaaaaa	ccagctcctg	gattcattga	ttttttgaaa	gggtttttgt	49680
tntctatct	ccttcatttc	tgcctcgatc	ttagttattt	cttgccttct	gctagctttt	49740
laaggtgttt	gctcttgctt	ctctagtctt	tttaatgggt	atgttagggg	gtcaatttta	49800
tttttccct	gttttctctt	gtgggcattt	agtgtctata	atttccctct	acacactact	49860
aatgtgt	cccaaagatt	ctgatattgt	gtgtctttgt	tctcgttggt	ttcaaagaac	49920
tttattt	ctgccttcat	tttcttatat	accagtagt	cattcaggag	caggttggtc	49980
cccatg	tagttgtgtg	gttttgctgt	agtttcttaa	tcttgagttc	tagtttgatt	50040
atogc	ctgagagaca	gtttgttgta	atttctgttc	ttttacattt	gctgaggagt	50100
tttttc	caactatgtg	gtcaattttg	gaatagggtg	gggtgtggtgc	tgagaagaat	50160
ttgt	ttgatcggg	gttttagagtt	ctgtagatgt	ctattagggtc	cacttggtgc	50220
ttgt	tcagttcctg	gatctgtctt	gttgatctgt	ctaataattga	cagtggggtg	50280
agagtcctc	ccagtatatt	tgtgtgggag	tctaagtcctc	tttgtaggtc	tctagggact	50340
gcttttatga	atctgggtgc	tcctgtattg	ggtgcatata	tatttaggat	agttagctct	50400
ttgtttgaa	ttgatccctt	tagcattata	tgatggcctt	ctttgtctct	tttgatcttt	50460
ttgttttaa	agtctgtttt	atcagagagt	tggattgcaa	accctgcctt	ttttgttttc	50520
ttgtgcttg	gtagatcttc	ctccatccct	ttattttgag	cctatgtgtg	tctctgcacg	50580
tgagatgggt	ttcctgaata	cagcacactg	atgggtcttg	actcgttata	caatttgcca	50640
gtctgtgtct	tttaattgga	gcatttagcc	catttccatt	taagggtaat	attgttatgt	50700
ttgaatttga	tectgtcatt	atgatgttag	ctgggtattt	tgtcgttag	ttgatgcagt	50760
ttcttcctag	cttcgatggg	ctttacaatt	tggcatgttt	ttgcagtggc	ttgtaccggt	50820
tgttcatctt	catgttcagt	gcttccttca	ggagctcctg	taagcaggcc	tggtagttag	50880
aaaatctgtc	agcatttgct	tgtctttaaa	ggattttatt	tctccttcac	ttatgaagct	50940
tagtttggct	ggatatgaaa	ctctgggttg	aaaattcttt	cctttaagaa	tgttgaatat	51000
tggccccac	tctcttctgg	tttttagagt	ttctgccaag	agatcagctg	ttagtctgat	51060
gggcttccct	ttgtgggtaa	cccagacctt	ctctctggct	gcccttaaca	ttttttccta	51120
catttcaact	ttggtgaatc	tgacaattat	gtgtcttgga	gttgctcttc	ttgagtagta	51180
tctttgtggc	attctctgta	tttctgaat	ttgaatgttg	gcctgccttg	ctaggttggtg	51240
gaacttctcc	tggataatat	cctgcagagt	gttttccaac	ttggttccat	tctccccgtc	51300
actttcaggt	acaccgtca	gacatagatt	tggctctttc	acatagtcct	atatttcttg	51360
gaggttctgt	tcgtttcctt	ttactctttt	ttctctaaac	ttctcttctg	gcttcatttc	51420
attcatttga	tcttcaatca	ctgataccct	ttcttccact	tgatcgaatc	ggctactgaa	51480
gctcatgcat	gcatacagta	gttttctgtc	catgggtttc	agctccatca	ggccatttaa	51540
ggtcttctcc	atgctgttta	ttctagttag	ccatttgtct	aatctttttt	caagggtttt	51600
agcttctttg	caatgggttc	gaacatcctc	ctttagctcg	gagaactttg	ttattaccca	51660
tcgtctgagg	cctacttctg	tcagcttggtc	aaagtctttc	tctgtctagc	tttgttccgt	51720
tgctgggtgag	gagctgtgtt	cctttggagg	agaagaggcg	ctctgaattt	tagaattttc	51780
agcttttctg	ctctgggttc	tccccatctt	tgtggcttta	tctaccttg	gtctttgatg	51840
atggtgacgt	acagatgggg	ttttgggtgtg	gatgtccttt	ctggttggtta	attttccttc	51900
taacactcag	gaccctcagc	tgcaggtctg	ttggagtttg	ctggaggtcc	actccagacc	51960
ctggttgccct	gggtatcacc	agcggaggct	gcagaacagc	aaatgttgca	gaacgacaaa	52020
tgttgctgtc	tgatccttcc	tctggaacct	tcgtctgaca	ggggtaccca	ggtatatgag	52080
gcgtcagtea	gcccgtatgg	ggaggtgtct	cccagttagg	ctacttgggg	atcagggacc	52140
cacttgagga	ggcagtcctgt	ccgttcgccc	atctcaaact	ccatgctggg	agaactacta	52200
ctctctttag	agctgtcaga	cagggacctt	taagtctgta	gaagttactg	ctgccttttg	52260
ttcagctatc	ccatgtcccc	agaggtggag	tctacagagg	cagtcaggcc	tccttgagct	52320
gtgggtggact	ccaccaggtt	caagcttccct	agctgctttg	tttaccact	caagcctcag	52380
caatggcaga	tgcccccgcc	tccagcctct	ttgcgcctt	gcagttcgat	ctcagactgc	52440
tgtgctagca	gtgagcgagg	ttccgttggg	atgggacctt	ctgagccagg	catgggatag	52500
aactctctgg	tgtgcccgttt	gctaagacca	ttggaaaagc	ccagtattag	gggtgggagtc	52560
tccctatttt	ccaggtacca	tatgtcacgg	cttcccttgg	ctaggagagg	gaattcccca	52620
acccttgcg	cttcccaggt	gaggcaatac	cccgccttcc	agctcacact	atgtgggctg	52680
caccactgt	ctgacaatcc	ccagttagat	gaaccagtt	cctcagttgg	aatgcagaa	52740
atcagctgtc	ttctgtgtcg	ctcacactgg	gagctgcaga	ctggagctgt	tcctatttgg	52800
ccatcttgga	accctgcctt	cttcattcat	atgtaataca	aaacttctaa	gggttttagtg	52860



gagaagagat	agagtaaaag	gatttctata	gacagaagaa	acagttgggtc	atcaactctt	52920
tcccttgtgg	cttcacatct	cccctaagga	cttcttatgt	tggtttgggtc	ttacagtata	52980
gtcagtgagg	gtttccctct	tcccattcct	tcctctgtct	gatttaaaat	gctgtttcaa	53040
gtatcgatat	aagtattttg	cctgtttctt	ttagcgtggc	tgtgaagggc	tgacattttc	53100
agaaggcact	tactgaaaaa	aaaaaaacaa	agaaatgtaa	gagtcacatca	catataaata	53160
gttaagtttc	taaaatatgt	at ttgagatc	ccagtaattc	tactaggata	aatagcaaaa	53220
attctccagc	cctgaagagt	tggctgtctt	ttcctttcct	tgttatcttg	attctctttt	53280
atttcatttc	acatgctaga	ccccatcatg	ttttcctgct	gcattcccca	ctccaccctc	53340
caaagcaatg	ttctcttctt	gcaacctgtt	tgaaaaaatt	gaactttgtc	ttatttcata	53400
tccctagtac	aagttgaata	tccctaatat	gaaatctgaa	atccaaaatg	ctctaaaatt	53460
cgactctttt	ttttttgaga	tgaagtcttc	ctctgtcact	caggctggag	tgcggtggcg	53520
caatcttggc	tcactgcaac	ctccgcagcc	tcccctaacc	ccatccctgt	gggttcaagc	53580
gattctccca	cttcacacctc	ccaagtagct	gggactacag	gcacctgcca	ccacacccaa	53640
ctaatttttt	gtatttttag	tagagatgag	gtttcaccat	gttgggtcaag	ctagtctgag	53700
actcctgacc	ttaagtgate	cacctgcctt	ggcctcccaa	agtgttggga	ttacagtcat	53760
gagccactgt	gcctagccaa	atatccaaaa	ctttttgagc	gcgacatga	tgctcaaagg	53820
aaatactgga	gtattttgca	ttttggattt	ttgggttaag	gatgctgaac	cagtaagtat	53880
aatgtaaaca	ttgcaaaatg	caaaaaaatg	taaaaaccct	aagcagttct	ggtcccaagc	53940
atttcgcata	agggatactc	aacccataat	cttttctttt	ctgttttctg	gttgggaagg	54000
catattggct	ttatagctaa	tttacacagc	attgatgttt	aatacacagt	gagtcceaag	54060
agatcactca	gacctattag	tagtttatta	gtgtcactca	cttctggaac	attctgtgat	54120
gttttacttg	gatgagttct	ttcacctctc	ttgggaaata	gtcataccaa	agtctgctta	54180
ttactacaat	gtgttatcta	ttatacattg	tccaattttc	tactcaaaat	tactagacag	54240
gcaaagaaaa	agtaaagtat	aacccttact	caagaaaaaa	aagcaatcag	tagaaactgt	54300
gagtgggccc	agatatggga	tttagcagac	aaagacttca	aagcagctat	tataatatgt	54360
ttaaagaatt	gaaagaaaat	atggtatcag	ttaaacagga	aatctaagta	gatgatataa	54420
actagacaat	aaaaataata	ttctaagttt	gaaaagtgta	gttactgaaa	ttaaaaattt	54480
acagacaaca	gcctcaacag	cagattagag	atagcaaaag	aaagattcag	tgactttgaa	54540
tgcaaggtctg	tagaaattac	taaattgatg	actctcacgt	agcaactttc	accctgtagt	54600
tggtttcata	tacaatgctt	taactctttt	tgctattttt	tctcttactt	tctgtgatgg	54660
aaacattttg	tgctttgatt	ttaatgggtg	tatcacagtt	atatacatct	ttacaaaactc	54720
atcaaattgt	atactttatt	tatatattta	tttttttttg	agataggggtc	tgactctttc	54780
tcccaggctg	gagtacagct	gaaattagtt	gggactacag	gtggggacca	ctgtgccagc	54840
taattttttg	at tttttgtg	gagatggggt	ttcccatgtt	tgccagctct	ggctcctcaac	54900
tccctgggtc	aatcattttc	gccgcctcag	actcccagag	tgttgggatt	atagggtgga	54960
gctactgtgc	ctggctggta	tagctttttt	tttttttttg	ggctcactgc	tgacagagtc	55020
tccctctgtc	gcccagggtg	gagtgccagc	gcgcgatctg	agctgggact	aacctctgcc	55080
tcccgggttc	aacccattct	cctgcctcag	cctcctgagt	agagtttcac	acagatgcgt	55140
gccaccatac	ccagctgatt	tttgatttct	tagtagaaac	agagtttcac	catgttggcc	55200
aggatgatct	cgatctcctg	acctcgatcc	acctgccttg	gcctcccaca	gtgctggcat	55260
tacaggcctg	agccaccatg	cctggctgtt	ggtatacttt	aaatggatgt	aattcattgc	55320
agattatacc	tcagtaaatt	tttatttagt	ttttgagact	gagttgctct	gttgcccagg	55380
ctggagtgc	gtggcacgat	ttcagctcac	tgcaacctct	gcctcctggg	tttaagcgat	55440
tttcctgtct	cagcctcccc	ggtagctgga	attagagggtg	tgtgccacca	tgatcagatt	55500
at tttttgtg	tttttagtaaa	gatgggggtt	caccatgttg	gccaggctgg	tctcgaactc	55560
ctgacttcag	gtgatccacc	cgctcggcc	tctcaaagtg	ctgggattac	aggtgtgagc	55620
caccacgcct	ggccttaatt	tttaataact	gtaaggctta	taaagaaaag	aatattcccc	55680
ttctgtttct	ttcctctcac	gtagcaacct	tcacccctag	tttggtttca	tatacagtgt	55740
tttaactctt	tttgctattt	ttttctctta	ctttctacta	tattttccaaa	tacaatgctt	55800
ctataatgat	tccttttttt	tttctatcag	tttttgataa	tcattgactc	cttatgggtca	55860
aagaagactt	aattcccttc	tgctactctt	catatatata	tataactaat	atataatattg	55920
tttagttttg	atataactaa	aaataactat	atgtatatat	aactatatat	gtataagcta	55980
tctgtatata	gttatatatg	tatatgcaca	tacatatata	gttacatatg	tatacacaca	56040
tacatatata	gttacatatg	tatatgcata	tacatatata	gttacatatc	tatatgcata	56100
tacatatata	gttacatatg	tatatgcata	cacatatata	gttacatatg	tatatgcata	56160
tacatatata	gttacatatg	tatatgtata	tgatatataca	tatatagtta	tatatgtatt	56220
agttatatca	aaataactat	gtatatatag	ctgtaaatgt	atatataaac	tatatgtata	56280
cagttatatata	tgtattttgta	tatgtgtgta	tacatatata	gttttttttg	tttttttttt	56340
ttgttttttt	tgtttttttt	tggagatgga	gtcttgcctt	gtcccccagg	ctggaatgca	56400



.

.

.

.

gtggtgctat	cttggettac	cgcaacctct	gcttcccagg	ttcaagcaat	tctcctgctt	56460
cagtctcccg	agtacctggg	attacaggca	cgtggcacca	cgccaggcta	atttttttgta	56520
tcttttagtag	aatgggggtt	tcaccatggt	ggccaggctg	ttctcaaact	cctgacctcg	56580
tgatccaccc	gcctcggcct	cccaaagtgc	tgggattaca	ggcgtgagca	ccgcgcctgg	56640
tccctataca	tatatagtta	tatataccta	tagttatfff	tagttatatc	aaaataacta	56700
tatgtatatg	taactatata	tatagtatat	atagtatata	tatactatat	agtgtgtata	56760
tatagtatat	atatactata	tagtgtgtat	atatagtata	tatatagtgt	atatatcgta	56820
tatacactat	atactatata	gtgtatatat	agtatatgta	gtatatatag	tatatatagg	56880
atagtatata	tagtatatat	agtgtatata	tactgtatat	atagggtgtac	atagtatact	56940
atatagtata	catatagtac	actgtatagt	atataatagta	tagtatatat	agtatacata	57000
gtatactata	tatagtatag	tatacatagt	atactatata	gtatatagag	tatatataca	57060
gtatactata	tagtatatag	agtatatata	cagtatacta	tatcgtgtgt	atagagtata	57120
tatacagtat	actatatcgt	gtgtatagag	tatatataca	gtatactata	tcgtgtgtat	57180
agagtatata	tacagtatac	tatatcgtgt	gtatagagta	tatatacagt	atactatatc	57240
gtgtgtatag	agtatatata	cagtatacta	tatcgtgtgt	atagagtata	tatacagtat	57300
actatacgt	gtgtatagag	tatatataca	gtatactata	tagtgtgtat	agagtatata	57360
tacagtatac	tatatagtgt	gtatagagta	tatacacagt	atactatata	gtgtgtatag	57420
agtatatata	cagtatacta	tatagtgtgt	atagcgtata	tatacagtat	actatatagt	57480
gtgtatagcg	tatatataca	gtatactata	tagtgtgtat	agcgtatata	tacagtatac	57540
tgtatagtgt	gtatagcgta	tatacacagt	atactgtata	gtgtgtatag	cgtatatata	57600
gtatactgta	tagcgtatag	agtatatata	tagtatactg	tatagtgtat	agagtatata	57660
tatagtatac	tgtatagtat	atagagtata	tatagttaca	tatacatata	gttatttttta	57720
gtaatatcaa	aaaaccatag	ctaagatfff	tatgatttag	caaaatattg	tttaccctcaa	57780
agccacgtat	tatatattaa	tgatattfff	tcccttttga	gtttttcttc	cccacagggt	57840
ttctgatctt	tttttttggt	ctctaactff	atattatagg	tatatactct	tttctctggt	57900
taatttgctt	tttaactaga	agacctcctt	tagtagttct	tttagtgaa	atctgtggtt	57960
aaactcttag	tctttccagt	gtctgaaata	ccctcattff	aatgtgacag	gtcacttttc	58020
tcacctcaac	atttttgaca	tacctcattt	ccaacaaaca	tttggtactt	ctgatagaga	58080
gtctgctgtc	attctcatac	ttgttccttg	tggaaatgtc	tatatataaa	gtctcagaac	58140
tatatatfff	taatataat	acatatfff	atttatffaa	aaatgtataa	ttaagttagt	58200
agtagtagta	gtagtagtag	tagtagtagt	agtagtagtt	tttttttttt	tagatggagc	58260
ctccatctcc	cgggttcaag	tgattctcct	gcctcagcct	cctgagtagc	tgggattaca	58320
ggcacacacc	accatgcctg	gctaattfff	gtatttttag	tggagatggg	gtttcaccat	58380
gttggccagg	ctgatcttaa	actcctgacc	tcagatggac	cacctgcttt	ggcctcccaa	58440
agcgctagga	ttacaggcat	gaaccactgc	gtctggcctt	aagttattat	tgactattga	58500
tgccctgttg	ctctatcaag	tagtaggtct	tatttattct	agaacctcgt	tttaagggtat	58560
cctgaatgtg	gctttttttt	ttttttttcc	taagacggag	tctcactctg	ttgccagggc	58620
tggagtgcaa	tggagtgtac	ttggctcacc	gcaatctctg	cctccttggt	tcaagcaatt	58680
ctcctgcctc	agcctcctga	gtagctgaga	ttacaggcgc	acgccaccac	acctggctaa	58740
tttttggtgt	tttagtagag	acgggggttc	accacattgg	tcaggcttgt	ctcaaactcc	58800
tgacctgtgt	atccgcccac	cttggcctcc	gaaagtctct	ggattacagg	cgtaagccac	58860
cacgcccagc	ctaagtgtga	ttttgttttt	ttttttaact	tgaattttat	ttatacttcc	58920
tgattttata	atagttaact	tgtagtaaaa	gtaaactggt	tgtctaaata	aataagctga	58980
tttcctatac	tagtttagag	tcttttagttt	tcttaaaccg	tattaagaga	tttgattaca	59040
aaagtaaaat	tgaggctttt	gtcctttttt	ttggccactc	taatttgctg	actgacttaa	59100
catttgctct	cacagtaaac	attgtaaata	aacactgatg	tttgttgctg	agtatgtaaa	59160
aatattatff	aggagtgtgt	agatttaaac	atttacaac	ctgtaatata	tatagtacag	59220
caacactgga	caaataacct	tttcaactta	atcagagggt	ttctataaat	gttcattttg	59280
atatatacag	cccatataat	acttagcata	tgtaaaagca	catctttata	ctctttcaat	59340
taaaattgat	tattattatt	attttttttt	tgagacagag	tccctctcag	tcacccaggc	59400
tggattgcag	tgggtgtgac	agggctcact	gcaatcgggt	agctggaaact	acagggatgt	59460
gccaccacgc	ccagctatff	ttgtatfftt	agtggagatg	gggtttcacc	atgttggcca	59520
ggctggtctt	gaactccata	cctcagggtga	tctgcccgc	tcagcctccc	aaatagctgg	59580
gataacaggc	gtgacttact	gcacccggcc	tagttttttaa	tagttttttaa	gacctctttg	59640
gtgggaacag	actttcatag	tgtttatagt	tataatagtt	tggaaatcaaa	gtagtcgagt	59700
gctacttttt	tttctctgtt	acatcctatc	cctaattgcct	ttactatatt	ttgtttctta	59760
tttgagagct	ttatttgctt	cataactfff	tttttttttt	tttttttttg	agacagagtt	59820
tcactctgtc	accagggtg	gagtgcagta	gtgcaatctc	ggcttactgc	aacctccgcc	59880
tcctgagttc	aagtgattct	cctgcctcag	cctccgaagt	agctgggatt	ataggccccc	59940



accaccacgt	ctgactgatt	ttttgtat	ttagcagaga	cgaggtttca	ccatgttggc	60000
cagcctgggc	ttgaactact	gacatcaggt	gatccgtcca	cctcgatctc	ccaaagtgtc	60060
gggattacag	gtgtgagcca	tcgcgcccg	cgcttcataa	cctttttata	ttgtctttct	60120
gttccccaaa	atgaagtagt	ttcttcttaa	aaatactgtt	ttttgttttt	ttgttttttt	60180
ttttatttga	gatggaatct	ctctctgtcg	ccaggctgga	gtgcagtggc	acgatctcgg	60240
ctcacttcaa	cctctgactc	cgtggttcaa	gcaattctcc	tgcctcagcc	tcccaggtag	60300
ctgggattac	aggcacaat	actgttttta	ataagaaaat	tagaataggg	acaggcacaa	60360
tggctcactc	ttgtaatccc	agcactttgg	gaggccaagg	tgggaggatt	gcctgagccc	60420
aggagttcga	gaccagccta	ggcaacatag	tgagtgaagc	cctgtttcta	taaaaaaaat	60480
ttttaattag	tccggtctgg	tggtgcacaa	ctgatgtccc	agctacttgg	gagtttgagg	60540
tgggaggatg	gatcacttga	gccaggaag	attaaggcca	caatgagctg	tgatcatgcc	60600
actgcaactcc	agcccaggcc	acaggtttca	aaaaagaga	aaattagaat	tatacgttag	60660
ggggaaaaaa	cctaataacc	atattcctat	ctacctagt	accaccccta	tgaagatcct	60720
ggtacatgtt	ctcattgaaa	tcttattttg	aaatattttt	gaattaaaaa	aatatattat	60780
taccttggtt	ttacaaatgt	ttgcttgtct	ttatgttcag	ctacaatttt	ctaatacaca	60840
tggttttact	ttttccttgt	cagaaaagg	agctgttcca	aaaagaaaag	accctttaac	60900
acacactagt	aattcactgc	ctcgttcaaa	aacagttatg	aaaactggat	ctgcaggcct	60960
ttcaggccac	catagagcac	ctagttagag	tggtttatcc	atggtttctg	gagtgaacaa	61020
gggatctggt	cctgtctcta	ccactcataa	ggtattctgg	gacagtaact	ttaattgctg	61080
tctttttgca	aatagaaaaa	tttttaagat	actattcctg	cttaagttag	tcataagtag	61140
tttataatac	tttagagaat	ggataagttt	ccataaagtt	aaattttttt	tttttttttt	61200
tgagactctt	gttttccagg	ctggagtgc	atggcgcgat	ctcggctcat	tgcaacctcc	61260
gcctcccagg	ttcaagcgat	tctcctgect	cagcctccca	agtagctggg	attacaggca	61320
tgcaccacca	cgtcgagcta	attttgtatt	tttagtagag	acgggggttc	tccatgttgg	61380
tcaggctggt	ctcaaacctc	cgacctcagg	tgatccgcct	gcctcagcct	cccaagtagc	61440
tgggattaca	ggcgtgagcc	accgtgccca	aagttaattt	ttttttaaat	ccacagggca	61500
acttgctata	aaactttttt	tggacctcat	tagacattta	ataccaagtt	ttgcttacct	61560
agtctgtttt	atataatgct	gtatatattca	ttttaaat	attttttaaa	ttatttcaat	61620
ctttgtaaat	gttttatacc	agtaacatac	ctagagtttt	gtcatgattc	taacaagggg	61680
taaaatttgt	ttttattttg	taaaaactaa	atctctggat	aaaatcctct	ataactgact	61740
taatgccaaa	tatagttctc	tagtgaatac	agttttacct	tcaggtaaat	aaatatacaa	61800
tttatggatt	tttataacct	ttttcctatt	tttaaagctt	gaattctgtg	aactttaagg	61860
tttaacttatt	tatgaaaagt	gtaaatgtta	ggttgtattt	tcataattaa	attttgtatc	61920
ctttaagggt	actccgaaaa	caaataaggac	aaataaacct	tctaccctca	caactgctac	61980
tcgtaagaaa	aaagacttga	agaatttttag	gaatgtggac	agcaaccttg	ctaactttat	62040
aatgaatgaa	attgtggaca	agtaagtttt	gccatctaaa	tgtttttatt	tatagttttt	62100
atattttta	tttacttata	aaacatgtca	ggagtgaat	agataataaa	taccttgtct	62160
ggtttacata	cttaattttt	atgatatgtt	tcaattataa	atgtagaaaa	ccattgcttt	62220
actgatttaa	ctatagttta	acagttaaca	ttaaaaataa	ctatatgtca	tagggcttag	62280
gcttcatctt	gtaataactg	ggcctgtttt	gtatcgtaga	actaactgag	gtcttgtttc	62340
ttagtggaac	agctgttaaa	tttgatgata	tagctgggtc	agacttggca	aaacaagcat	62400
tgcaagaaat	tggtattctt	ccttctctga	ggcctgaggt	aagaacttta	tattatcatt	62460
tttctataat	accatctgtt	actgaatcca	tagtagtagt	agtagtaaa	aaatatttga	62520
gctatactaa	aataattaat	tcataataag	taacaataga	tttaatgttt	taaaaaagat	62580
aaaacattaa	caattgatgt	tgagaagtta	ttatagaata	ggaaaatgga	tatgagttcc	62640
taattttcaa	tccttttata	tttttaggaa	aagtgggtag	tatgatataa	ttttgttttg	62700
ttttggagac	ggagtctcac	tcatttgtgc	gccaggctg	gagtgcagtg	gcatgatctt	62760
agctcactgc	aacctctgcg	tcccagaatt	caagtgatcc	tcctgcctca	gcctcccaag	62820
tagctgggat	tacaggcatg	tgccactatg	cccagcta	ttttgtta	ttagtagaga	62880
tgggggtttt	ctatattggc	caggctgggc	tttcatacct	gatgtcaagt	tatccaccca	62940
tttcggcctc	ccagagtgtt	aggattacaa	gtgtgagtca	ctacatctgg	ccaaattttg	63000
atatcaaggt	gagagagatt	taaaattaa	ataaggtaca	aaaattagcc	tagtgtgggg	63060
gcgcagctt	gtagtcccaa	ctactgggga	ggctgtggca	ggagaattgc	ttgaaccagg	63120
aggcagaggt	tgcatgtgac	caagatggca	ccactgcact	ccatcctggg	tgacagagcg	63180
agatgtcatc	tcaaaaacaa	aacaggccgg	gtacggtagt	tcacgcctgt	aatcccagca	63240
ctttgggagg	ctgaggcggg	cagatcacga	ggtcaagaga	ttgagaccat	cctggcta	63300
atggtgaaac	cctgtcttta	ctaaaattac	aaaaattagc	tgggcttggg	ggtgtgtaac	63360
cccagctact	cgggatgctg	aggcaggaga	atcgcttgaa	cccgggaggc	ggaggttgca	63420
gtgagccgag	atcacgccac	tgcactccag	cctggcgaca	gagcgagact	ccgtctcaaa	63480



aaaataaata	aaaatttaaa	aagataaata	cataaaaaata	aataaataat	attaagagaa	63540
ggaaatcagg	caggtagtgg	cccctgacac	aatgagtttt	cccagaattg	gattgcttgg	63600
aaatgccgct	caaagagtgt	ggtaaactcc	atcgaaggct	aaataccaac	gtgacagtga	63660
taataaacia	gtactttatg	ggaaagtttt	ttttttaatt	atttttaaaa	agagagaaat	63720
tgtactggag	aaaagaggaa	ttcaggtaga	ataattcttt	tttttttttt	ttttgagatg	63780
gaatttttgc	cctgttaccc	aggctggagt	gcagtggctt	gatgttggct	cactgcaacc	63840
tctgcctttg	gggttcaagt	gattctcatg	cctcatcctc	ctgagtagct	gggtctatag	63900
gcatgcaccc	caacacctgg	ctaatttttg	tgtttttagt	agagataggg	ttttaccatt	63960
ttggccaggc	tggtctcaaa	ctcctggctt	cagatgatcc	gccacacttg	gcttcccaaa	64020
gtgctgggat	tacaggcaga	ggccactgtg	cctggcatga	agaacaattt	ttaaaagaat	64080
gactttttaag	gatattaagt	catcaaagta	gatagagcca	ttaaatgatg	ggtagaaaacc	64140
taatcttcca	tcccatattt	tatgacttat	taaggaagat	aggcaatctt	gcagttgata	64200
ataaatatatt	ggctttccat	acttagcacc	gttttgaatt	tttccagttc	acagatggta	64260
tatttagtgc	tgttccaata	atattgcaaa	aaataaatct	gaagactcac	ttctgggtca	64320
cagttatttc	actattaaca	ttaaaatctt	acggacctac	ctgcaacatg	tagtggaat	64380
aagttgtgtg	gcacgttgtg	gtgggtgcat	tattaaataa	atgtgcaaag	gttttatggc	64440
tgtctcaatt	tttccatagt	cttgagtcac	tcacatactg	cctttatggg	gtttatcact	64500
tgtagttag	taccaactgt	actagaactt	actgaatatt	tttcaagtct	attttactta	64560
ttactcattc	taagtacata	aaatcactgg	tctgatatgc	aagttaggtt	ttcgctaata	64620
cccagtaaaa	taattagata	attattaaaa	tagaaagcat	ttgtcattat	gtgaactgga	64680
tttatgtcat	gtaccacctc	aggtcacact	ttgggaaacg	tagcttacgt	attgagctaa	64740
catttttgctg	gagttatata	agattagtaa	atgctataga	ctaaatattg	aaaaaaaaagc	64800
ttgcttaaa	aggatttaca	taattgtaga	cctgaaaagg	atcttagcag	tcttctgac	64860
gattcattga	taaagaaact	agtcctaaga	gctaaaatga	cttgctcaag	attttagtta	64920
agtggtagac	ctaggaata	cctttctaag	attagagctg	cctgggtagt	gctatagctg	64980
ccacctaatt	cttacattta	tttattaagg	gacatgcttt	ttacggtaga	aaaatatatc	65040
taaaattgtg	aaggtaatac	caaggaaata	tttaggctat	gcaagaaaag	tcattcagct	65100
tttgctttac	ccaggcctat	agagaacata	tctaaaataa	taagtaacat	gtgaatttga	65160
atthagatcat	attttcatgg	tagtgacata	atgtgcaatt	aagaagctat	aataatgtaa	65220
aaatgccagt	tgcaagttat	ttcaaataaa	gttgaataac	ctgacctcag	gaggggcaga	65280
aaccttagta	aacttagcaa	actgagacct	aaggaagagc	aagatctgtg	gtagagccag	65340
gtctaaagg	ttatatcctg	gtacaaactt	gtctatacat	caccacacct	gccaaatgga	65400
actttttttc	cttgtagttc	agatttccaa	gaaagggaat	attattgtcc	tagctcagcg	65460
tttcacaata	gactcacaaa	atgaatcttg	ttaagcctat	gaatgggtaa	gacttatccc	65520
ataacacgcc	tctattgtgt	aaaaaaatca	gctttattcc	taattccctg	tagtaggaag	65580
cttgcataga	ccataataga	gacattgtta	gattatatca	agaagggtgag	tggcaggcat	65640
tcttgagtat	actggaattt	ttatttcatt	gtagaaacac	agaatcataa	ttagtcatct	65700
aaaactgggt	ttaaggagat	aataactaaag	atactcactt	gaggtagatt	gttctcagaa	65760
ttacttaaaa	taagtgcctt	aaatcttgct	cattcatgaa	gccatagccc	tggaaggaga	65820
tattgaaaaa	cttctccatc	ttccttcaat	cccataggac	catgttttta	attgtagttt	65880
aacattaaac	ataatttttg	aaagtagcca	atattaatct	ttgacaaaaa	actcaaacta	65940
aatagcagaa	aatagaacat	gccctctcct	ctcgcccaaa	agtaagacta	catttttaaag	66000
tatatcacag	tgatctgtct	aaatatTTTT	ctcaaataat	ttactaagaa	agcgtatatt	66060
cttctgtcct	tatttcatag	atctcttctg	tatcctgtgc	agcttgtcta	tcatttccag	66120
cttttatatt	tcttattttg	ttctcttgct	gcttaaatgt	gattctttga	gatttataga	66180
agaatgtgat	tgttctccag	ttttaaaaag	ctcttctcta	tctactaaaa	tttccatgaa	66240
tttctttctt	tctttttttt	tttttgattg	agacagtctc	gctgtgttct	ccagcgtggc	66300
cttaggetta	agtgatcctc	ctgtctcagc	ctcccaagta	gctgagatta	cagaaatttc	66360
catgaatttc	caaatatgat	gttatctttc	atataagctt	aaacataagc	ctttcttctt	66420
tcaaaaatgt	tagaacttac	ctctttaaat	aatcttataa	agccatttct	taatttttgt	66480
taccttttgc	aatctattaa	cttgcatatc	atagtaaagt	aacccttagc	tccatcatac	66540
cacacacctt	ctactcctgt	cctcccaact	tttctcacc	aggctggagt	acagtgaata	66600
tttagtgctt	ttaaatatata	ttcattacca	agccaagtag	cgtactatatt	tttcttctga	66660
cagccttttt	ttccctctaa	taattgtgtt	attttttcat	ttgcttcatt	ttcatcttat	66720
cgattgataa	ttcttctggt	ccatttccagt	ctaatttcca	catggccaaa	ccaattttaa	66780
aacctgttaa	ctcaggtcct	tttttttttt	ttttttccca	gaataacctt	tcctggaacc	66840
accttcttct	tttctatcga	gaccaaattgc	tttccatgcc	tgtgtcaaa	tggtgtgtct	66900
ctaactctgtc	ttcactacca	tcataatttc	cttttctctt	tttcttctct	ttattatgta	66960
tctcctgtct	tttaaatctc	atattttcca	ctttcatgat	ttattttctc	attttgatgg	67020



.

.

.

.

.

ctacatcttt	cagtaactga	gaaaagggtg	tgtgagggaa	tattttgaga	tgctgtagtt	67080
ttgaaaatat	ctttattctg	ctgtcttctc	attaataatt	tgactg atg	tcaagattta	67140
agttggaac	tattttccat	gagtattttg	aaggcattat	ctattgtctt	ctagcttcca	67200
gtgctgcttt	gtattagaga	cagaagatca	ggaatcagga	tagcattgga	cttcttattt	67260
ctatttccca	aagctttttg	gatcttctct	ttattcttat	cttctgatat	tttataatga	67320
tatgcttgat	gcatttactt	tttcttttct	tcttccctct	tgtacatttc	tttctttttt	67380
taggtggggg	caggggggaa	ggagtctcac	tcttctcacc	caggctagag	tgcatgtgtg	67440
caacctcagc	tactgcaac	ctccacctct	ggggttcaag	cgattctcct	gcctcagcct	67500
cccagtagtc	tgggattaca	ggcatctgcc	accacgcca	gctaattttt	gtatttttag	67560
tagagacggg	gtttcaggct	ggtcttgaac	tcctgacctc	aggtgatccg	ccctcctcag	67620
cctcccaaag	tgctgggatt	agaggcatga	gccactgcgc	ccagcccctc	ttggaccttt	67680
ctagatattc	atccttcaat	tctgagaaaag	gttctataat	tctttaattt	aaaaaaagtt	67740
tttttgtttc	atctttctca	cttttttttt	tttttttttt	ttttaagaca	cagtcttgct	67800
ctgtcaccca	ggctggagtg	cagtgggtgcg	atctcagctc	actgcgccct	tcgcctctctg	67860
ggttcgagtg	attctcctgc	ctcagcctcc	caagcagctg	ggattacagg	tgccaccac	67920
caccatgcc	ggctaattgt	tgtattttta	gtagagacgg	agtttcacca	tgttggccag	67980
gctggtttca	aactcctgac	ctcaggtgat	ccacctgctt	tggcctccca	gagtgtctggg	68040
attacaggca	tgagccaccg	tgcccagccc	tattctgatt	tcatagatgc	agtgtctttt	68100
atgtctctga	gaggttttgt	caagttttct	tttgtttact	ttattatctg	tttcttcaag	68160
attcctcctc	ttttcttttt	gccttaagct	ttttcatatt	ggaagcttct	ctcaaagtgc	68220
tggttattct	ggttgtccac	ttatattacg	tggaaacaata	aaaaagttga	ttaggactct	68280
gtgctgtaag	taagtaaatt	gttgatagtg	agagcccatt	aatcacagga	tgatcaagca	68340
gcaagcctac	tgtttatttg	ggtttctcca	aatggctata	gctgtcagtc	ttttttttct	68400
ggggttattt	ggcttctcta	gaaaagaatt	ttctagtctc	ttgtttgag	aatacaagct	68460
tgactattgg	tgttccagaa	aatgggtggg	aggagacgac	tttgttttct	gttaggttga	68520
gccatgtgaa	attactaggt	tttgttttgt	tttctgtttt	gttttgagac	agagtctcac	68580
tctgttgccc	aggctggagt	gccttagcat	gatcttggtc	cactgcaacc	tctgcctccc	68640
aggttcaagc	aattctcatg	cctcagcctc	ccaagtagct	gggattatag	gcatgcacca	68700
ccacgcctgg	ctaatttttg	tatttttagt	agagacgggg	ttttgccatg	ttgaccagtc	68760
tagtctcgaa	ctcctgacct	caggtgatcc	acctgccttg	gcrtcccaaa	gtgctgggat	68820
tacaggtatg	agccaccaca	cctgggctga	aattactgtt	tttatagggtc	aaaaacagtt	68880
gagggatagg	catggtggct	cccacctata	atcccagcac	tttcagaggc	caaggcagga	68940
agattgcttg	agtccaggag	ttcgacacca	gcctgggcaa	tgtagtgaaa	ccccatctat	69000
atttaaaaaa	aaaaaataga	tgaatatctt	tgtttgca	taacatgatt	gcatatgtag	69060
aaaatcccaa	agaaccaaca	aaaagagctc	ctagaactaa	taagtgatta	tgacaaggtg	69120
tagaatacaa	agttaatata	caaagtacaca	ttgtttttt	atctaccagc	agtgaacaac	69180
tggaatttga	aattaaaaca	caatactgct	gggtgcagtg	gctcacacct	gtaatcccag	69240
cactttggga	ggccgaggtg	ggcagatcac	ctgatgtcag	gggttcaaga	ccagcctgac	69300
caatatggtg	aaaccccatc	tctactaaaa	atacaaaaat	tagctgggca	tggtggcggg	69360
tgctgtaat	ctcagctact	caggaggctg	aggcaggaga	atcgcttgaa	cccaggaggc	69420
agaggttgca	gtgagccaag	atcgcgccat	tgactccag	cctgggggac	agagcaagac	69480
tccgtctcaa	aaaaaaaaaac	acaatacctt	tcatattaac	actaataaaa	tgaaatatgt	69540
gtagttctaa	caaagtttgt	tgtagaagat	ctatatgaga	agaattatag	cactcatgaa	69600
agaaatcaaa	gatctaagta	aactgagaga	tattccatgt	aaatggacag	ggagactaaa	69660
tattattgag	atgtcagttc	ttcccaagtt	catatatcga	ttcagtgcag	tcccagctaa	69720
acccagcca	gttattttgt	ggatactggc	aaactaaagt	ttatatgaaa	aggcaaaaga	69780
cctagaacag	ccaacacagt	attgaagaag	aaaaaagtca	gaggactgaa	actaccaat	69840
ttcaagactt	actgtaaagc	tacattaatc	aagacagcat	gtcattggca	aaagaataga	69900
caaataaatc	agtagaattg	gacagagagc	ctagaaatca	acccacacag	ataaagtcaa	69960
ctgatctttg	gcaaagggac	aaagacaatt	cagtggagaa	aagatagcct	tttcaacaaa	70020
tggtatagga	caactggaca	tccacatgca	aaaaagttaa	tctagacaca	gacctgacaa	70080
ctttcacaaa	aataaatgga	tcatagacct	aaatgtaaca	tgcgaaactga	aacttctaga	70140
ggataacata	ggagaaaatc	taggataaca	tgagaaaaat	ttttggtttg	gcagtgactt	70200
cttaggtaca	ataccaaaac	atgatccttg	aaaaaaaaaa	tcagtatgtt	gaactttgtt	70260
taaattataa	acttctgctc	tgtgtaagat	gctgttagga	gaatgaaaag	acatgcagca	70320
gagtgggaga	ctttacaaaa	ttcattatct	gatgaaggac	cagtatccaa	aatatacaaa	70380
gaacttttga	acttcaacaa	taagaaaata	tacaaccag	ttaataaatg	ggcaaaatat	70440
ctgaacgagc	acctcaccaa	ggaagataga	tagatgacaa	caagcatatg	aatatatgct	70500
caacatcatg	tcgttaggga	aattgcgcct	taaaacaaca	acaagatacc	ctgccatccc	70560



tattagaatg	gctgaaatct	aaaacactga	caacacccaaa	ttctggcagg	gatgtggagc	70620
agcagaaact	ttaattcatt	gctgatggaa	atctaaaatg	gtagaaccat	tttgggaagg	70680
agttggacag	tttttttacag	aactaaagac	agtttgacag	tttcttataa	aactcttacc	70740
atatggtcca	gcagtcctac	tccttagtat	ttacccaaat	aagtttaaaa	tgtacatcca	70800
ataaaaaaaa	ctgcacatga	atatttctag	cagcattatt	catagttgcc	aaaacttgga	70860
agcagtcagg	gcacccataa	gtaggggaat	ggataaacag	actttggtat	atcatgtaat	70920
ggagtattat	tcagcaataa	aaagaaatga	tctatcaagc	cacaaaaata	tatggaggaa	70980
ccataaatgc	atattgctaa	atgaaagaag	ccagtcctgaa	gaggctacac	tataggattc	71040
tgactatatg	atgttttgga	aaaggcaaaa	ctatggaaac	agtaaataga	tcagtggttg	71100
ccaagggaga	cagggagaga	tgaataggtg	gagcacagtg	gatttttaag	gcagtgaaac	71160
tgttctttat	gataatccaa	tggtggatac	atgtcattat	acctttgtca	aaaccacacg	71220
aataaaaaac	ataagagtga	accctaattg	aaaatatgga	cttcagttaa	taataatata	71280
tgaatatctt	ttcattagtt	ctaacaagtg	tactacacta	atacaagata	ttcagagtag	71340
gggaaatttg	aaaggaatga	gaggttatat	gggaactctg	tactttctgc	tcaattttct	71400
gtaaacctaa	aatcactaaa	aaaaagttt	atttttattt	ttattttttt	ttatttttta	71460
attttttgag	atggagtttc	actcttggtg	cccaggctgg	aatacaatgg	cacgatctcg	71520
gctcactgaa	acctccgcct	ccagggttca	agcgattctc	atgcctcaac	ctcccagagta	71580
gctgagatta	caggcatgtg	ccaccacgcc	cagctaattt	tgcagtttta	gtagagacag	71640
ggtttctcca	tggtggtcag	gctggtcctg	aactcctgat	ctcaggggat	ctgccgcct	71700
cggcctccca	aagtgcctg	attataggcg	tgagcagctg	cgcccagcag	gttttttttt	71760
ttaagttgga	tattagccat	ttcatatgat	tcaacttaaa	agtacataca	ccttcactgt	71820
tattaaagct	taggtgagat	gttttcagtc	tggagctcta	cccttgattc	ctgccatgcc	71880
tagtgtccct	gaatctggag	actctgactt	atttctttag	agaacgaac	tctgccttc	71940
tgctttggtg	gtgattggtg	cctgcttgac	tgcctttggt	ggggagttcc	tcataccaac	72000
tttcaatcag	ttcctgtgtt	tgtttggttg	tttgtttttg	agacggagtt	ttgctctgtc	72060
gcccgggttg	gagtgcagtg	gcgcaatctt	ggctcaatac	acgctccgcc	tcccgggttc	72120
acgccatttt	cctgcctcag	cctcctgagt	agctgggacg	acaggcaccc	gccaccacgc	72180
ccggctaatt	tttttgattt	tttagtagag	acgggggttc	atcgtgttag	ccaggatggt	72240
ctcgatctcc	tgacctcgtg	atccgccgcg	cttgccctcc	caaagtgcctg	ggattacagg	72300
cgtgacgact	gtaagccacc	gtgccgggcc	agttcctctg	tttttgactg	cctgccttac	72360
tactgctttc	tgtggtgcct	gataacatcc	aattcctgaa	ccttcctggg	atttttgttc	72420
acgtcagcgt	gcttcttgca	tttaggtatc	ccttcacaag	taggcattta	ggttttaagc	72480
tctgctaagt	gatttaccac	actttatcag	ttctccattt	tgtggaattc	attgtgttaa	72540
tctccttttc	tatttttggt	tggaagattc	atataatttt	tattcattta	gtattttttg	72600
gtgggatttc	taagtagaga	aacatgccta	tgttcaatat	gtcttgttta	agcagtcctgc	72660
tttcattttt	accaccgagg	agttgggttt	atttcttttt	tttttttttt	tttttttttt	72720
ttttgagaca	gggtctcact	ttgtcagcca	ggctggagta	cagtggcaca	atcaaagctc	72780
actgcaccct	tggcctccca	agctcaagtg	atcctcccac	ctcagcctcc	tgaatagctg	72840
ggagctcagg	tgatccagca	cacccagcta	attttttttt	tttttaattt	ttttagaag	72900
cagagttccc	ctatgttgcc	caagctggtc	ttgaactctt	gggtcctagt	aatcctcctg	72960
tctttgcccc	ctaaagtctg	ggattacagg	tgtgagccac	caacccagc	tggttttaat	73020
tattacatta	atgttgattt	atgtgttttc	tttaaagcta	tgggcagctc	tggttgggaa	73080
gatgctactg	aaaaaaggat	gctttttaga	tggcaaagag	tacttaaaat	gtctctagaa	73140
tcatagttgt	aaactaaagt	atataatttt	tagttgttca	cagggttag	agctcctgcc	73200
agagggtgt	tactctttgg	tccacctggg	aatgggaaga	caatgctggt	aagggtcttc	73260
ttcaaatttg	agttttctgt	tgagatattt	gggataatat	gaaaaaaaga	aactttatct	73320
tgctccttag	tctattattt	acgacttgct	ttttgctatt	gtacactttt	gttttttttg	73380
tttgtttggt	ttgttttggt	ttgagtgatc	tgggctcacc	gcaacctccg	cctcctgggt	73440
tcaagcgatt	ttcatgcctc	agcctcccaa	gtagctggga	ttacaggcac	acgccaccac	73500
acctgtctag	gttttttatt	tttggcagag	acaggatttc	gccatgttgg	ccagggtggt	73560
gtcgaactcc	tgacctcaag	cagtcctcct	acctgggcct	cctaaagtga	tgggattaca	73620
ggcatgagcc	actgcacctg	gctgatacac	ttttaagttt	ttcagctact	tttcaatgta	73680
gaagtagatg	gaaaaccatg	tacgttatct	tcagtagtgt	gtttttggtt	ggttaaattt	73740
gacagtatga	ttgtcattat	tttttgtaaa	ttaaattttt	acctggaaga	gcttacctta	73800
ctatattgag	tatctttcta	acctctgatt	tttgcttcta	ctatcataat	aactttatct	73860
aagtaatcag	tatgttatag	cttttttttt	tttttaagta	ttcttttgcc	agaagttttt	73920
atcaggctct	ggatacctct	ttcctctgca	tagtcctcct	ggatggaaga	aacaaagagg	73980
gaaagagtaa	cttttcctta	gatgtttgtc	tttctcaaag	cagttatctt	tgtatatcta	74040
agaagagagg	agaataacac	tgtctctctt	tttttttaaa	tctctctcta	ctcattctct	74100



ctcaggagga	gagtagaaa	aagcacagct	cttcctataa	cctgtcctta	ttactgagaa	74160
aggaacacat	tgattgccat	gtattgggga	ttgtattata	ccttacattt	ttatttttat	74220
tttttattaa	tttttttttt	tagacagtct	tactctatca	cccaagctgg	agtgcagtga	74280
tgtgatctcg	gctcactacc	atctcttcct	tctgggttca	agcaattctg	ccacgtcagc	74340
ttcctgagta	gctgggatta	caggcatgcg	ccaccacgcc	tgactaattt	ttgtattttt	74400
attagagatg	ggatttcacc	atgttgccca	ggctgggtct	aaactcttgg	cctcaagtga	74460
tgcgcctgcc	tcggcctccg	aaagtgtctg	gattacaggc	atgagccacc	acacctggcc	74520
tcatagctta	cattttttaga	gaatcttttc	tagtacttaa	atcggtaaat	atgggtatct	74580
tttaaagtga	atatattgaa	ctaattttaat	atttgctctt	gtgattttta	aaggctaaa	74640
agtagctgc	agaatcgaat	gcaaccttct	ttaatataag	tgctgcaagt	ttaacttcaa	74700
tacgtgag	tgctctgttt	ccaatattgt	cgtattttta	gttactgtct	aaatgttact	74760
taactg	taaatggtaa	tatttcatga	aaatattttt	ctaggagctt	atctattgta	74820
tattt	acatatgatg	aatatctatc	ttcagagtag	aaagttagtg	acatttgggt	74880
tac	tgtattagtt	tactggggcc	atgtaataaa	ataccgtaaa	ctggctggct	74940
	caattttattg	tctcacagtt	gtggaaggta	gaagtctaag	atcagtcaaa	75000
	gttgcttcc	tcctgagggc	tgtgagggaa	aatgtatgtt	gtgtgcctct	75060
	gtggctt	gctggcaatc	ttttgtattc	cttggcttgt	agatgcaccc	75120
	tttttate	tttatgtggc	attctccctg	tgtctgtcac	catgtccaaa	75180
	ggattaggac	ccaccctaatt	gatctcaatt	taagtttgtc	atcagcaaca	75240
tgctc	caaataatgt	cacattcata	ggtactaggg	ataggacttc	aacacgtttt	75300
aacac	agttcaaccc	attaacaaat	actatcactt	tccacttaag	cttcaagtaa	75360
gggatttt	atctcaagga	gccaccagat	aggaacacag	atctgatggc	ataaaactgag	75420
tttctggc	cttctgatcc	tgatctaaaa	tatgacagaa	gattttccct	gttttaattt	75480
tttttttt	tttataatag	agacacagtc	tcgctatggt	gccaggctac	tttcaaatcc	75540
ctgggcttaa	gcaatccctc	tgccctcgac	tctcaaagtg	ctggaattac	aagtgtgagc	75600
caccacacct	ggccagtaga	ttttccctgc	tttcttttga	ttgtttataa	ttttgttttc	75660
tttttccatt	caccctctgc	tgaccctata	gtattattca	aagaagtgtt	cagtctagtt	75720
ttggggtagg	gcaagcataa	ctacagtgtc	taaagagagt	aatttgtctg	gtgtgcagaa	75780
ctagtgtgta	aatataactg	gtgcattgca	aaactgtgaa	gtagtttctg	tcaaacctta	75840
cactgctttg	tctttctccc	tctctccctt	tctcttggtt	gccccctccc	cctcccaatg	75900
ataccttagt	ctctgcttgc	ctttatcaaa	acctttatga	ttggccgggc	acagtggccc	75960
atgcctataa	tcctagcatt	ttgggaggct	gaggcaggaa	gatcacctga	gcccagaagt	76020
tggagaccag	cctaggcaac	atggtgaaac	cccactctct	ccaaaaataa	aaaatagagc	76080
caggtgtggt	ggcatgcacc	tgtagtccca	gctactcagg	aagctgaggc	gagaggatct	76140
cttgagcctg	agaggttgag	gctgcagtga	actgtgatca	tgccactgca	ctccagcctg	76200
agtgaacatg	tgagaccctg	tctcaaagaa	acaaaacaag	ggggggcatg	gtggctaact	76260
cctgtaatcc	cagtactttg	ggagactgag	gcaggaggat	tgcttgaggc	caggagtcca	76320
ataccaccct	gggcaacata	gtgagacccc	catctctaca	aaaaataaaa	aatttagctg	76380
gacatgccag	cgaatacgtg	gtcccagcaa	atcaggaggc	tgaggtggga	ggatcacttg	76440
agtcaaggag	gttgaggttg	cagtgaacca	cgatcatgcc	aatgcattcc	aacctgggag	76500
tcagagcaag	accacgtctc	aaaaacaaaa	caaaacaaac	ttttatgggt	gaaagtgttt	76560
tggcaaacat	acttaaactg	aaatgtgaat	ctctgatgaa	agaacatgtt	acctgtaaaa	76620
gtttgaagtg	tcagcatttg	ttgcaccgaa	atccagaggt	gaggccagggt	gtggtggctc	76680
acgcctgtaa	tcccagcact	ttgggaagcc	aaggctggca	gatcacctga	ggtcaggagt	76740
tcaaaaccag	cctggccaac	attgcaaaac	cccgctctga	ctaaaaatac	aaaaattagc	76800
cagacttggt	ggtgtgcgcc	tgtaatccca	gctattcggg	aggctgagac	acgagaattg	76860
ctcgaaccga	ggaggcagag	gttgcagtga	gccgagatgg	caccactgca	ctccagcctg	76920
ggcaactgag	tgagactctg	tctcaaaaaa	aaaaaaaaaa	aaaaatccag	aggtgaatcc	76980
agaggtgatc	accacatgat	atccagatag	cctctttcca	tgagaggctc	aaaggataat	77040
tttactgtct	acagttttgc	agcgagagaa	acttgatttt	atcagtacac	caagagcaga	77100
tctatgtctt	cggaacagac	atgagatcag	aattgtctag	ctgctatgaa	cagcatgttc	77160
tctccctgta	cctatagaca	tgtatgggaa	acttattttg	aaggttgtat	aatgagcagt	77220
gagttaaagc	aaacttgaca	tgttgacccat	agttgttatg	gcattggact	aaagtagcct	77280
ccatcactat	gatagagata	gccttggtgt	tggaccataa	gtttttgaag	tctgtttgct	77340
aagactccct	cttccttagg	cagttcttca	aatattatta	tgctttttcc	tgtcttggga	77400
ttctcttcca	tgaggtttcc	tttctcagga	acactatctt	ctcctgggta	atttgtactt	77460
attcttcaga	taccagttga	aatgttacaa	ccttccctga	ccctccaaac	tattccccgt	77520
tattattctc	ctagcaccac	ttgcacctca	ttttcatact	cactggagtt	gcaattcata	77580
ttcattgata	ggattatttt	aattgtatct	gatatcactg	tcacctccac	tagaagatgg	77640



tctccatgcg	ggcagagact	atcaccatgt	gttcttcact	tcaattttca	gtagttggct	77700
gtgagtaggt	attgaataaa	tattttgtgga	gtaatcataa	tgaggatatag	atattattct	77760
catagctcta	ttttattaat	tagggaatta	cagaattcag	tgatctgctc	aggatctcat	77820
aaccaggaag	tgggaaacta	ggatttgagc	tccagttagt	gtggcctttc	attaaaaata	77880
ttacagcaac	cattctcttt	tttttttttt	ttttttttga	gatggagtct	tgctctggag	77940
tgcagtggcg	tgatctcggc	tactgcagc	ctctgcctcc	cgggttcaag	tgatccccct	78000
gccttagcct	gctgagtagc	tgggactaca	ggcaccgcgc	acgacacttg	gctaattttt	78060
gtatttttagt	agagacgggg	tttcaccatg	ttggccagga	tggtctcgat	ctcctgacct	78120
tgtgatccgc	ccgcatctac	ctcccaaagt	gctggattat	aggcgtaagc	caccacaccc	78180
ggcctacagc	aaccattctc	ttttatccat	acttttttca	agagtactgt	ttcatcttca	78240
tgttttcaga	aacaacatag	cattcatgat	cttaaccccc	aattctgata	ctgcctgaat	78300
atcttgaagt	aagtttactt	ttaagaaagt	tgaggctagg	tgtggtggct	catgcctgta	78360
atccagcac	tttgggaggc	caaggcaggt	ggatcacttg	agctcaggag	ttcaagacca	78420
ggctgggcaa	catggcgaaa	ccctgtctct	accagaaata	caaaaaatta	gtcgggcgtg	78480
gtggcgtgtg	cctgtggtcc	cagccacttg	ggagactgaa	gtgggaggat	ttcttgagct	78540
tgggaggtgg	agtttgctgt	gagccgagag	atcatgccac	tgtactccag	cctgggtggc	78600
agagtgagat	cccatctcaa	aaaaagaaaa	gaaaaagaaa	attgaaatgt	ctagtctatc	78660
attttgtcag	ttctatctaa	tacaattttt	tccttatgtc	taactgaaat	ctgctttttc	78720
taatttttac	atacttgatt	taacaaaact	caatcttttt	ttttttttat	gagacagcct	78780
ttcaaatata	taggaactta	atgttatatc	tgcttccctc	cagtccccag	aatagttact	78840
attttagttg	tctttctatg	gtctcatgcc	agtttgtcag	tatgcctaga	taagaactga	78900
atattttacc	tcagatgtga	cctgactttg	aagacttaaa	aaggaagcat	tgtgccaggc	78960
gcagtggctc	acacctgtga	tcccagcact	ttgggaggcc	gagtgaggca	tgctcaggat	79020
ttgagaccag	cctgattaac	atggagaaac	cccatgtttc	tctactaaat	acaaatctct	79080
actaaatact	aaatactaaa	tctctactaa	aaatacaaaa	gatgagctgg	gcattgtggc	79140
acatgcctgt	aatcccagct	actcgggagg	ctgaggcagg	agzatcactt	gagcccagga	79200
ggcggaggtt	gcggtgagcc	gagattgcgc	cattccactc	cagcctgggc	aacaaaagtg	79260
aaactccatc	tcaaaaaaaaa	aaagcagcag	catttgttaa	tattatgtag	atgttgtgtc	79320
tcatgatcta	tcttgagaaa	gcttttgga	ggaaactgat	catagtcatg	gacaacattt	79380
gtgttattaa	aatatctaga	ttattttcca	caaaaaatca	gttacatatg	tatcttaaca	79440
tgttgtatta	ttgtttaacc	ttgtttattg	aataactaac	atgtagaaaa	gtattttatg	79500
cataagtata	cagctgtata	cagctcagtg	gattaccaca	aagcgaatat	actttcataa	79560
tcaccacca	ggtcaagaaa	taaattgtta	cctgtggccc	taaaatccct	ccagccactc	79620
caccatcttt	atccactcac	tctctccct	caaaaccact	agactactaa	catcatagac	79680
aaagctagca	tgcttttgaa	ctttatataa	atctaattgat	gtaggatttt	gtgtgtatgt	79740
gtattttggct	tctttcatca	gcattgtatt	tgtgagattt	atccagattg	ttgcaagtag	79800
ttgtagtgtg	gcttttttac	acagatttaa	tttttatatt	tttcttatat	gttcgaacag	79860
ttaacctgct	tatctattat	taaaaaaaaa	aaaacgaaca	ttcacatagt	tcttaccagt	79920
ttacagtgtt	ttttccacac	cgtcttcaaa	atgtaaagt	tggtcttcaa	tacatcagta	79980
tgctgtctaga	tttaaatact	agggaaaaaa	aaatcagaga	agttaataat	attaactgtc	80040
acctccacta	gaagatggtc	tccatgtgga	cagtaatat	tctctgtgat	tatctgtgtc	80100
aagtcaaatc	ttctgtaaat	ttctttaaat	attttaataa	atcatagtac	ttaaaatcgt	80160
ctcaatatct	taaagtagtt	aaaagtaact	ataaaaatagt	acctgttttt	ctgatcacat	80220
tttacttcct	atgtgaaatt	ttacaagtcg	ttactctatt	tatttattga	tttatttttt	80280
aagacaggg	ctgttctgtc	gcccaggctg	gagtacagtg	gcgtgatcat	ggctcactgc	80340
agcctcagcc	tcttgggctc	aagtgatcct	cccaccttag	catccaagt	agctgggact	80400
ataggcacat	gccaccatgc	ccagctaatt	ttaaaaaatt	ctggggggcc	gaatgcggtg	80460
gctcacacct	gtaatcccag	cactttggga	ggccgaggca	ggcgaatcac	aaggtcagga	80520
gttcgagacc	agcctggcta	acatggtgaa	accctgtctc	tactaaaaat	acaaaaaatt	80580
agcgggggtg	ggtggcaggc	gcctgtaate	ccacttactc	aggaggctga	gagcaggagag	80640
ttgcttgaa	ctgggagcca	gagattgcag	tgagccgaga	ctccatctca	aaaaaaaata	80700
attttttttt	gtagtgcaca	ggtgtcactg	tgttgccagg	gctgggtctca	aacttctggg	80760
ctcaagtgat	cctccatttt	cggcctccca	aagtgtcagg	atcacaggca	tgagtcaactg	80820
tgcttgggtc	tcaagttggt	attaaagcat	gtttaccac	attatgcaca	tggtataatg	80880
gaaagtattg	ttgtggaagt	taggagatag	ggattctagc	ctagcttttt	attttttttg	80940
gacaaggtct	cactttttcg	ccccaggccg	aagtgcagtt	gtgcgatctc	ggctcactgc	81000
aacctccaac	tctcaggttc	aagcaattct	cccacgtcag	cctcccaggt	agctgggatt	81060
acaggcatgc	gccaccacgc	ccggctaatt	ttttagtatt	tagtagacac	agggtttcac	81120
catgttgggc	aggctggtct	tgaactcccg	acctcaggtg	atccacccac	cttggcctcc	81180



caaagtgcgtg	ggattacagg	catgagccac	cgcacccggc	ctctagcgta	actttttacat	81240
cctgaactga	ccttaagaaa	gtataacttt	aggcctgttt	catctgtaaa	atgttaatgt	81300
cataggagat	gatcttttga	gatttctttc	agctctgata	attttgtgtg	tgtgtgtgtg	81360
tgtgtgtgtg	tgtgagatgg	agtcttgctc	tgtcgccgg	gctggagtgc	agtggtagca	81420
tctcggctca	ctgcaagctc	cacctctctg	gttcacgcca	ttctcctgcc	tcagcctccc	81480
gagtagctgg	gactacaggc	gcctgccacc	acgcctggct	aattttttgt	attttttagta	81540
gagacgaggt	ttcaccgtgt	tagccaggat	ggtctcgatc	tcctgacctc	gtgatccgcc	81600
cgcctcagcc	tcccaaagtg	ctgggattac	aggctctgagc	cgccgcgccc	ggcctagaat	81660
atttgtttcg	attattctaa	atctggtgac	atttcttttg	tttttaagtt	aaatcttcag	81720
taaaaagaat	aaatgccacc	tagaggacag	aaaaattttt	acagtagatt	atcacagacc	81780
tcatgactca	ttactttggg	gtataaaatg	gcctttgtat	ggtgtcagca	cctgggaatg	81840
tctcaagggg	tgttcattga	ccttctggac	tatctggaaa	tattttgata	tttattggct	81900
gggcgcgggtg	gctcatgcct	gtaattccag	cacctcgga	ggccaagtta	ggtggatcac	81960
ttgaggtcag	gagtttgaga	tcagcctggc	caactaagtt	agttgactat	ttgtgaaatt	82020
tccctctcca	aataggaaag	ataggattgg	aaacattatt	cagaagggaag	aagttttaaa	82080
gaagggcagg	cttaaagact	atctaataaa	tttaccata	tggtaaaacc	ccatctttac	82140
taaaaataca	aaaattagcc	aggcgtgggtg	gtgggcgcct	gtaatcccag	ctacctggga	82200
tggttagggga	tgagaatcac	ttgaaccag	gaggcaaaag	tcacagttag	ccgagattgc	82260
accactgcac	tccagactgg	gctatagagc	aaaactctgt	ctcaaaaaaa	aaaaggaaaa	82320
aaaaaaaaaga	aattttgata	tttatgtgag	aatgactttt	cacggtgttc	ttaatagcgc	82380
aagttttgtt	taggagagca	cattccaact	tacttgcttc	tataaatata	ccgtgtaatc	82440
tagggcttaa	agaatatgta	caatgtcttt	ttctctcccc	tagtcttccc	cttttctcac	82500
tagttatata	cttgactgaa	gagctatttc	attctcaagt	cttaggaatg	caggggtgaag	82560
caaaacagat	gaatttttaa	agcacttaac	caggctgtat	gaaatcacag	tctgttgtct	82620
aaaattgtaa	gggacgggta	gtagtactct	cccctttctc	aaaccaaatc	tttggttggt	82680
ttaaggaagg	gaaattaaat	tcctgtgtgc	tagattttca	acataaaatt	taaaaaactg	82740
gaataatgtt	gcattttatg	tgtataacag	tataatgctt	tgtttttaggt	gggagaagga	82800
gagaaattgg	tgagggctct	ttttgctgtg	gctcgagaac	ttcaaccttc	tataattttt	82860
ataggtaaga	acataatttc	caactaagtt	attgactatt	tgtgaaattt	ccctctccaa	82920
ataggaaaga	tacgatttga	aacattattc	agaagggaag	agtttttaag	aagggaagc	82980
ttaaagacta	tctaatgaat	ttagtaggac	ccactatatt	aataagtagt	aaactagatt	83040
aatctcagat	gactcacata	gcttggctct	taattaaagt	cttatacttg	tatttctctc	83100
agatgaagtt	gatagccttt	tgtgtgaaag	aagagaaggg	gagcacgatg	ctagtagacg	83160
cctaaaaact	gaattttctaa	tagaatttga	tggtgtaagt	gttgattatg	atatttttaa	83220
tgtggcagca	tttttagtata	ttttcctatt	aatggccaa	ggttaaaaaat	acaaatatct	83280
ttatatttgt	tattactttt	ctaaatgaat	tgaaaaaaga	ttttttgctt	gtaggtacag	83340
tctgctggag	atgacagagt	acttgtaatg	ggtgcaacta	ataggccaca	agagcttgat	83400
gaggctgttc	tcaggtaggg	agattttatat	ggaaatacat	gcatttatta	cagacaatat	83460
ttactactgt	gtccacttta	catattattt	ccttactctc	agttttaaga	ctaaattcac	83520
tattttcttc	cagtagctac	tctagcctct	tgttaccaac	tacataagga	ttttgaggtc	83580
aacagcttga	tatcaggaga	attagtctag	taaagggtta	gttacgggtt	taggcaaat	83640
ggtcagttct	tctgggttgc	attaaattat	ctttaaactt	gaaaattgat	tcttaccgtc	83700
ttttattggg	cttcatgaga	aaataatcaa	atcagggttc	atggaaaatg	tgttttctgt	83760
tcttgtgtgt	ttgagcagtt	tattcccttc	atactggaag	gacgggttgg	gtcagcataa	83820
aattcttggg	catcctttcc	caaggaattt	acaatattgc	tccattgttt	ctagattggc	83880
tgtagtagtg	taaaagtctg	aatccaaatt	gattttttca	aatttgtaaa	ataacttgat	83940
ttttttctct	ctttactgga	tggtttacgt	gggtatattg	cacccaggta	gtgagcatag	84000
ggtgcagtag	gtagtttttc	aaccacacc	tcactgcttt	cttcccccat	ttagtagtcc	84060
gcagtgtgta	ttgttcccat	gtttatgtcc	atgtgtactc	aatttttagc	tccccactta	84120
tgagaacatg	agctatttgg	tttactgttc	ccacattaat	tggcgtagta	tttaaggctc	84180
cacctccatc	catgttgctg	caaaggacat	tatttcatcc	ttttttatgg	ctgcgtagta	84240
catagtattc	catggtgtgt	acgtaccaca	ttttctttat	ccagtccacc	attgatggac	84300
acttagattg	attctatgtc	tgctgtcatg	aatagtacag	caataaacat	gaaatgcag	84360
tcttttttgg	ataatgatct	attttccctt	gggtatatac	ccagtaatgg	gatttcaggg	84420
tcaaaggtag	atttgtttta	tggtttttga	gaaatcttca	aactgcttcc	tacaatgact	84480
gaactttaca	ttcctaccag	cagtacatac	gctccacaac	ctcaccaaca	tctgttattt	84540
ttttactttt	tcccgccaat	ctgtagacaa	tatgggattt	tttttgccat	tttattaata	84600
gccattctga	ctggtgtgag	atatctcatt	gtgattttga	tttgcatttt	tctgtggtt	84660
agtgatgatg	agcatttttt	catgtttgtt	gtccctcgt	atgtcttttg	agaagtgtct	84720



attcatttcc	tttttaccca	ttttttaata	gggttggtg	tttttagctt	gttcaattgt	84780
ttgaattcct	tatagattct	ggatatcaga	ccattggttag	atgcagttct	gtagattgtc	84840
tgtttactct	gttgatgggt	ctttttgctg	tgacagaagct	cttttagttta	attaggtccc	84900
acttgatcaat	ttttattttt	gttgcaattg	tggttgggga	cttagccaaa	atttctttct	84960
gaggtgatg	tcaagaagt	tatttcctag	ttttcttcta	gcatttttat	actttgagat	85020
cttacattta	aatctttaat	ccaccttgag	ttaatttttg	tatatggtga	aaagtaagg	85080
tccagtttca	tttttctaca	tatggctagc	cagtaatccc	catgccattt	attgaatagg	85140
gagtcctttt	cctcattgct	tgtttttgtc	agtcttatca	aagatcagat	ggttatagg	85200
gtgtggcttt	atttctcaat	tttctgtcct	gttcccttgg	tttgtgtgtc	tgtttttata	85260
ccagaattat	gctgtttggg	ctgctgtgta	gttcgggtta	aagtcaggta	acgtgatact	85320
tacagctttg	ttctttttgc	ttaggattgc	tttgtctatt	tggactcttt	ttttgcttcc	85380
atatgaattt	tagaatattt	ttttctaatt	ttgtgaaaaa	caacattgat	agtttgatag	85440
aaataccatt	gaatctgtaa	attgctttgg	gcagtatagt	catttttacga	ccagataaat	85500
caggcaagag	aaggaaacaa	aaggcatcca	aataggaaaa	gaagtcacat	tactctcact	85560
cttcactggc	aatatgattc	tatactttga	aaaccctaaa	aattccgtca	aaaggctact	85620
agaactgata	aatgatttta	ccccagatt	caggatacaa	aatcagtgta	caaaaaatta	85680
gtaccatttc	tatacgctaa	taatgtccag	gctaagagtc	aaatcaagaa	cacagtccca	85740
tttacaatag	ccacaaagaa	aatgaaataa	tggctgggca	cagtggctca	tgccgtgta	85800
cccagcactt	tgaggagctg	aggcaggtgg	atcacctgag	gtcaggaggt	cgaggcaggc	85860
ctgaccaaca	tggtgaaacc	ccgtctctac	taaaaataga	aaaaattagc	caggcggtgt	85920
ggtggggcgc	tgtaatccca	tctgctcggt	aggctgaggc	aggagaattg	cttgaaccca	85980
ggaggcatag	gttgacgtga	gccgagatcg	tgccactgca	ctgcagcctg	ggtgtgaaag	86040
agcaagaatc	cgtttccaaa	attaaaaaaa	aaaaaaagaa	aaaaaaaaga	caatggagta	86100
cctaggaata	cggctgatga	aggaggtgaa	agaatctgta	tgaggagaag	tgtaaagcac	86160
tgctaaaata	aatcagaggt	gacacaaaata	aatggaaaaa	cactccatgc	tcatggattg	86220
gaagaatcaa	tatcgttaaa	atgggaaatt	tatcttaatt	tttttcccta	gttcattttt	86280
ctgttttctt	ttgcacataa	attatgtgtt	atgttgcgct	cttttggttt	ccctatctgt	86340
aattttcttc	ttcatccttt	ttaactctta	ggtctttttg	ttttgttctg	ttttcccaag	86400
cctgttgctt	ttatttctta	ctgggttagt	attctaactc	ctttctgcta	tatttctgtg	86460
aactcttctt	tccttggtgg	ttttttatct	attactttcc	tgacatctgt	tattctactt	86520
ttcatcttct	gtttttatttc	tttgagcatc	tgatgtttat	ttccataagc	tcttacattt	86580
cactgggctc	ttgtttcata	gaaaacaatt	ttttggcaag	gacagaccac	ttgaggccag	86640
gagtttgaga	ccagcctggc	caacatagta	caaccctgtc	tcigctaaaa	atacaaaaaa	86700
ttagctgggt	gtggtcacac	acacctgtaa	tcccagctac	tcaggaggct	gataatcgct	86760
tgaaccagag	aggcagaggt	tgacgtgagc	caagacggcg	gcactgtact	ctagcctggg	86820
ggacagagca	aggctctgtc	tcaaaaaaaa	aaagaaaaaa	ttttttgggt	aggcgctcat	86880
gtgggtcaca	cctgtaatcc	cagcactttg	ggaggccaga	tcacttgagg	ccgggaattc	86940
cagaccagcc	tcgccaacgt	ggcaaaaacc	catatctata	aaaaaaactt	aaggataaaa	87000
aaaattaaat	ttttaaaaaa	ttaaattatt	ttatttctgt	gaaagatact	tggtcataat	87060
tttagtttgc	tctatggcaa	aatcttctca	tgcatgttct	tcattttatc	tagtttatgt	87120
ttattgtgtt	ttgtttgtct	ttttttgggg	gcagggtttg	ttgttggtgt	tttttttttt	87180
tttttttttaa	atgagccctg	tcctgggttt	tccttttttg	cttattattt	ttctttgaat	87240
aaaaggaggt	ttcttggaag	acctagtttt	ggaaggatat	tgtaggggaa	gggataggta	87300
tcttgagtag	tccatgtata	cactaaacta	gcattgcagc	tccttcactt	acagtgaagc	87360
cctaccta	aatgaagtat	atgtgtatgt	aaataagaga	taggggtgat	ttattcttct	87420
actaaaaaat	attggctcact	gcaagacatt	gtcttcccgc	cagactgtta	gatgccttct	87480
tataaagata	attcgtctta	gagtttcttt	gttcattctc	accttccttg	ccattctctat	87540
gctatcaatt	agggtttata	tggtctgcag	agccaggctt	catatgcttg	tttaaatgtg	87600
gagggactgg	ccttattacc	ctttagggtg	tgctatctct	ttttgggagg	aaaactgtac	87660
tctgacagtc	atgtccatga	tccttagcgg	cagatccttt	gtgtgtcttc	ctttgtcagt	87720
gtaattttca	ctgccctagg	cagttcacct	ttatgtattt	tagttccata	aatgttacct	87780
agtttggtga	tgaagtgtat	gtttctgttt	ctattatcct	tgctgatttt	agtttttttc	87840
agattcaaga	agacagatct	acttatatca	gatattttgt	ttgttttaaa	gcattactgc	87900
tattaaaaac	atatacatat	ataaatgatt	aattctaaca	ataccgaatt	aaggatttta	87960
aacattaact	agagtttttaa	aagagtaaca	aaaatatttt	tacattgata	actacaaaaa	88020
tgcatctcag	tgccctgaat	attatatttt	aaaagctttt	cctgtcattt	gctgtttcag	88080
ctttaaattc	aaaattatat	ttctaaaagt	gctggatttt	tttttttagg	cgtttcatca	88140
aacgggtata	tgtgtcttta	ccaaatgagg	aggtagtat	ctgtgtttga	attttttttg	88200
tttttagagca	gaaacaagaa	ctaccatctt	gacaatatta	agtcttccaa	tccatggtac	88260



agctactttg	gaaaacagtt	tagtggtttc	ttaaaaagat	gtacataagc	ttactgtcag	88320
agccagcaat	ttcactccta	agaattttacc	ccagagaaac	aaaaatgtat	gtccacacaa	88380
agacttgtag	aagaattttt	atagcagcaa	tattaataat	agccagaact	acaaatgatc	88440
taaaatgttt	attttggtga	ataaacaat	gtggtatatt	catttaatgg	atactgttat	88500
ccattataca	tgctgcatac	tcacttcata	tattaagttt	tatgaatgaa	actccaaaac	88560
attataagtg	aaagaagcaa	gatgatatat	gtcgcataac	ataggattct	gtttatatgc	88620
agaaaacgca	aattttat	ttattttatt	atttttttga	gatggagtct	tgctctgtcg	88680
ctcaggctgg	agtgcagtgg	cgtgatctcg	gctcactgca	accttcacct	cctgggttca	88740
agtgattctc	ctgcctcagc	ctcctgagta	gctgggactg	caggcatgtg	ccaccaggcc	88800
tggctaattt	tttgtgtttt	tagtagagac	gggtttcact	gtgttagcca	ggatgggtg	88860
atccgcctc	ctcagcctcc	caaagtgtcg	ggattacagg	catgagtcac	cgtgcccgcc	88920
cagaaatcga	aaatttctag	aaacgtaaag	cagatcagtg	ggtgttcagg	gtgagagtgt	88980
ggaatgagga	ttaactgcaa	acaggtatga	gtgaactaaa	aagtgttcta	aaactggatt	89040
gtggtgatga	ttgcacatct	ataaatgtat	taaaactcat	tgaattatat	acttacaatg	89100
ggtgaaattc	atagtagatg	aattcatacc	tctataaaac	tggttttttg	cagcaaaata	89160
tttattgcta	tggtttttatt	caagtgtggt	gaattattgc	tatgttttta	tccaggtgtg	89220
gtgtctcacg	cctataatcc	caacactttg	ggaggctgag	gtgggaggat	cgcttgagcc	89280
aggaattcaa	gaccagcctg	ggcaacatag	ggagaccctg	tctctacaaa	aactttttta	89340
gaaattaacc	aggcatggtg	gcacgtgcct	ttggtcccag	ttactcggga	ggctgagatg	89400
ggaggattgc	ttgaaccag	gatatcgagg	ctatagttag	ctatgattgt	gccctgcact	89460
ccagcctggg	taacagcaca	agaccctgtc	tcaatataaa	aaaagaaaag	aatcattaat	89520
tctgaaatta	gactgaatga	tcattttttta	atattttttca	gacaagacta	cttttgctta	89580
aaaatctggt	atgtaaacia	ggaagtccat	tgacccaaaa	agaactagca	caacttgcta	89640
ggtgagtaat	ttggatttgg	tttatcttac	agctttttatt	tattttttgt	aaataattct	89700
tttttttagtt	atttaaagta	atcttaagta	tgaaatgagt	aattcattga	tcagaagact	89760
ttctctcatc	ctctacctcc	tttgcctttt	tcttacctct	tgttcttata	tatatatgtg	89820
gaatctaaat	ttcactggct	atgtcctttg	caagacgtga	tctaatgatg	atagtatatt	89880
ctttttataa	atttattaaa	acctgtaagt	ggtattaaaag	taatttaaac	atttacacct	89940
taggcaacat	ggtaaaaccc	catctctaca	aaaaataaaa	tgagtcagga	atggtggcac	90000
atacctgtag	tcccagctac	ttgggaggtt	gatgtgggag	gatcgcttca	gcccgggagg	90060
tggaggttga	gcaaccagat	gagaccctgt	ctcaaaaaaa	atttttaaca	tttcatattg	90120
cacactaaca	cagatatttt	tatataaaact	tggtcttggc	atatgctgaa	tacttttcag	90180
ctcttttaat	aatttttaatt	atttgataat	tttaataatt	ttaatttcag	ctatttgaat	90240
aatcttaaac	tttatgaaca	gtgtgaagta	gaaactgagt	ataggagatg	cgctcatgaa	90300
ttttattccc	tgaaaaatgt	cataaagtaa	tataaagtat	tcaatgagta	tgtgagtatt	90360
gaataatttt	aaacaagaact	ctgagttccc	agatacctat	gacaaaagaa	caaattactt	90420
ggaagtctct	tctactccag	acctttctgt	ctatccccct	atttaaaaaa	atgtacattg	90480
attgtccctt	atcccatcaa	gagagtgttc	tttgccaaga	taaacgctaa	ttagagaagt	90540
aattgtgttt	atgtgtcagt	tggccaccag	tagttttagt	atgttacctt	tccaggtgac	90600
atgtttacaa	tttaatgaag	cctctggaga	tagtatgcct	taatctaaag	gaatgggtct	90660
aatgattaaa	attttaggga	ttactacatt	tgttttcagc	aaatgggcct	ttctattccc	90720
ttaaaatcaa	acctcataaa	ctatgggttt	gctgacaaaag	ggaaactagt	ttttactctt	90780
atttatatgga	gatttttaaat	ctgcagagat	gaaaataatt	tgctgtttca	gctgggtgag	90840
gtggctcatg	cctgtaatct	cactactttg	ggaggctgtg	gcgggtggat	cacctgaggt	90900
caggagtctg	agaccagcct	gaccaacatg	gcgaaacgcc	atctatacta	aatacaaaaa	90960
aaaaattagc	caggtgcggt	ggcgcatgcc	tataatccca	gctacctgag	aggctgaggt	91020
aggagaatcg	ctccaggagg	tggagatcac	agtgaagctga	gatcatgcca	ttgactcca	91080
gcctgggcaa	caagagcaaa	actccatctc	aaaaaaaagc	gggaggggaa	ataatttgct	91140
gtttcttctc	tcccttctct	agaatgactg	atggatactc	aggaagtgac	ctaacagctt	91200
tggcaaaaga	tgagcactg	ggctctatcc	gaggtaggta	tacaagagct	taaaacattt	91260
agaactattt	attataccac	cttagaagtt	taagaagtcc	aaaaaaatct	accaagagat	91320
tttttttttc	ttttggagac	agggtctcac	tctgttgccc	aggctgtagt	gcagtggcac	91380
aatcatagtt	cactgcagcc	tcgaaccccc	caggctcagg	tgattctccc	accagcctt	91440
cccagtagct	gggactacag	gcatgcacca	tcatgcccag	ataattattt	tattttattt	91500
cttgtagaga	cagggtttca	ccatgttgcc	caagctggtc	ttgaactccc	gggctcaagt	91560
gagcagcctg	ccttggcctc	ccaaagtgtc	gggattacag	gcatgagcca	ccatgccag	91620
ccttaaaacc	attcttagct	cacagatcat	acaaaaacag	tctatgagcc	agtagactgt	91680
gaccctgtt	ctaggcaaga	ggtttcattt	cttgatagta	agtaagcagc	agaaaactca	91740
aaaagaaaac	tttaaacaaa	aaagaataat	gtaaggaaaa	tgttcttctc	tgttttctgg	91800



ctctttttcag	tattctgaca	tcatttgaaa	ttaatgtgtc	taaaggaaat	tctataagag	91860
aagttttcac	ttttttaatt	atatcttttg	agggattgca	gcaaaatatt	tattgctatg	91920
tttttattca	ggttacttgg	ttgattctga	aataccacga	ccaattcttt	ttaggttagg	91980
ttaggataaaa	ctcacacaga	tattttcagc	tacatttcca	cagccagcat	cggtagtggg	92040
tggttgatgc	tctcaggtat	gcactcagat	tattagctat	gatgataaag	taatcataat	92100
agctccatcc	ctttctctca	ttcaagaaag	tacaattgaa	tgtagttgta	cgagaaagat	92160
gttatcatag	tgtctactgt	aatttgcatt	aaaagcctac	attatacaaa	cctttttgtt	92220
ttactagtta	aaaattacca	gtaacacctc	ccgtagttaa	atagggtgct	gattaagaac	92280
tgctactttg	caaaaataaaa	aatgtaatct	tgagagtata	ttttgaaact	ctggatgaac	92340
aaattttctt	gctgctggag	cttaaatctt	gttcagttag	tggatttaga	aacagtagct	92400
aatgcataaa	tgagttgtat	tttcttagaa	cgttttggct	attctaaggt	agacaaaatt	92460
tgggtttaaa	aaaagattcg	catccattaa	agtacagaaa	atggccgggt	gcagtggctc	92520
acacctataa	tcccagcact	ttggggagcc	aaggtgggag	gatcgtgtgg	gcctaatagt	92580
ttcagactag	actgggcaac	ataggggagt	cctgtctcta	caaaaaaata	taataaaatta	92640
accaggcccc	ttggcatgag	cctgtaatcc	cagctactcg	ggagtctgaa	gtaggaggat	92700
cacttgagcc	tggaaggtgc	aggcagcagt	gagccatggt	catgccactc	actgcactcc	92760
agcctgggtg	ggtgacatag	tgaggtgctg	tcttaaaaag	taaataaata	ggccaggcac	92820
ggtggctcac	gcctgtaatc	ccagcacttt	gggaggccaa	gauggggtgga	tcatgaggtc	92880
aggagttcga	gaccagcttg	gccaatatgg	tgaaccgcga	tctctactga	aaatacaaaa	92940
attagccagg	cgtggttgtg	cgcacctgta	atcccagcta	ctcaggaggc	tgaggcagga	93000
gtatcgcttg	aacctgggag	gtgaggttgc	agtaagccga	gcactgcagc	ctgggcaaca	93060
gagcaagact	ctgtctcaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaagtaa	atacataaag	93120
tgacagagtaa	aaacaaaaaa	agactaatgc	attttgtaaa	gaacaagttg	cattctttta	93180
agttcccttt	caaatttgtg	aactatttgt	tttgggcagt	atgcaagaaa	ttgaacactt	93240
tccagttatt	ccaaagaagg	aataattctct	tctcagcatt	tataaattgt	atttgccttc	93300
aaagttaaca	tgtgtctctt	ttttttaata	taatgatttg	tactgaatag	atacatgtag	93360
atcattgtac	ttggttttgc	ccttcaacaa	tttcaactgc	aaaatgtatg	tattttttaag	93420
tgctgactt	ttatgtttta	cagaactaaa	accagaacag	gtgaagaata	tgtctgccag	93480
tgaggtatag	tatttttacia	tgatattttc	tttgtcttct	atattgtaag	acatatataa	93540
gacatacata	tatgaatgtg	tgtgtgtgtg	gttttttttt	tttttttttt	ttttttgaga	93600
cagtctggca	tgatatcggc	tcaactgcaac	ctccacctcc	tggattcaag	ctattctcct	93660
gcctcagcct	cccaagtagc	tacaggcatg	taccaccaag	cccagctaag	ttttgtattt	93720
ttagtagaga	gggcatttca	ccatgttggc	cagtctggcc	tcagactcct	gacctcaagt	93780
gatccaccca	actcggcctc	ccaaagtgtc	gggattacag	gcttgagccg	ctgcatctgg	93840
ctgtttttgtt	tgtttgtttg	tttgtttgtt	tttgttgagac	ggagtctcac	tctgtcgccc	93900
agactggagt	gcagtgggtc	gatctcagct	cactgcaagc	tctgcctccc	gggttcacgc	93960
catttctctg	cctcagcctc	cctagtagct	gggactacag	gcgcccacca	ccacgcctgg	94020
ctaatttttt	gtacttttag	tagagacggg	gtttcaccgt	gttagccagg	atggtctctg	94080
tctctgatc	tcgtgatctg	cctgcctcag	cctcccaaag	tgctgggatt	acaggcgtga	94140
gccactgcgc	ccggcctttt	tttttttttt	tttttgagac	acagggtctt	gctctgttct	94200
caggctggag	tgacgtggtg	caatcgtggc	tgactgtagc	cttgacctct	tgggtctcaag	94260
cagtcctctt	gtctcaaaact	tttttagtagc	tgggactaca	ggcaagtgcc	acaacgcccc	94320
actaattttt	taattttctg	tagagatggt	gtctctccct	gttgcccagg	gtggtctcag	94380
actcttggcc	tcaagtaatc	ctcctgcctc	agtctctcag	agtgtttggg	actaattttt	94440
tttaatggat	ttttctagtt	gattagaagc	tagaaaaaat	taactttgct	tcattttcaaa	94500
atatggaaac	ctatgaaata	gtcaaaaattt	tttttggttaa	tagacaaaaac	atatatacat	94560
aaagtcattt	tctatcaaat	catattacat	aaataccaaa	cagaagaaaa	ttacaggccg	94620
ggtgcggtgg	ctcacgcctg	taatcccagc	actttgggag	gccgaggtgg	gcagatcact	94680
tgaggtcagg	agttcaagac	cagcctggcc	aacatggtga	aaccctgtct	ctatcaaaaa	94740
tacaaaaaaa	ttaggccggg	ggcgggtggc	cacacctgta	atcccagcac	tttgggaggc	94800
cgaggcgggc	ggatcacgag	gtcaggagat	cgagaccatg	ctgggctaaca	tgggtgaaacc	94860
ctgtctccac	taaaaaatac	aaaaaaatta	ccggggcggtg	gtggcggggca	cctgtagtcc	94920
cagctacttg	ggaggctgag	gcaggagaat	ggcgtgaacc	tgggaagtgg	agcttgacgt	94980
gagccgagat	cgtgccactg	cactccagcc	tgggcgacag	agcaagactc	catctcaaaa	95040
aaaaaaaaaaa	atagctgggc	atggtggcac	gtgcctgtgg	tcctggctac	tgggaagct	95100
gagccaggag	aatcacttga	acttgggagg	cagaggttgc	agtgaagcaa	gattgccccca	95160
ttgcactcta	gcctgggcga	cagagtgaga	cccagttctca	aaaaaaaaaaa	aattacatga	95220
aacttgtggg	ggtttataca	ataaaaaata	cctataacta	tacgttaatc	ctttaataat	95280
gttctatttt	gtttttcaat	aacctttatt	tttttaatta	atttcttttt	tttgagaagg	95340



actctggctc	tgtcaccag	gctgcagtc	agtggcacca	tcttggctca	ctgcagcctc	95400
cacctccag	gttcaagcga	ttctcctgcc	tcagcctcct	gcgtagctga	gattacaggc	95460
atgggccacc	acgccagct	aatttttttt	ttttttgtat	tttttagtaga	gatgggggtt	95520
cactatgttg	gccaggtggg	tctcgaactc	ctgacctcag	gcgatcgcca	gcctcggcct	95580
ccctaagtgc	aaggattaca	ggcgtgagcc	accatgcctg	gccaatttat	taccgtttct	95640
taatatggaa	gacctcctta	tgagatgctg	aagcatttca	tttgaaaaaa	gttgcatgta	95700
aaatataaaa	ttgggtatgc	agaatgggtca	taactctgtg	agcaaaattt	tgaaatcagg	95760
cttgaattac	agtcagcctt	ttctatccgt	ggattctgca	ttcatggatt	caaccaactg	95820
cagatcaaaa	atacttgggg	gaaaaaagca	tctgtgttga	acatgtacag	actttttattc	95880
ttgttattat	tccctaaaca	atatagtata	aaaactattt	acattgcatt	tacattgtat	95940
taggtgttat	aagtaatcta	gagatgattt	aaagtataca	ggaggattgt	tgaggttgta	96000
tgcaaatact	gcaccatttt	acataaggga	ctttaagcgt	ttgcggattt	tgctatctat	96060
gggggaacca	atttccgaga	gatactgagg	gacagctgta	tatttghtaac	ttatttttta	96120
tttcccta	tgacagcagct	gttgagggga	cagtgaactg	ttaacacaga	taacaagtgt	96180
attcaagtac	attttggagt	ttgtttgttt	ttttttttca	ggctttttatt	cgctgtcgg	96240
atgaggcacc	atacttgaat	tttttttttt	aagaaagctt	ttagttttct	ttctttcttt	96300
tttttttttt	gtgagatgga	gtcttgctct	gtttctaggc	tgagtgagg	tgacgtgggtg	96360
tgatctcagc	tcactgcaac	ctccgcctac	tggttccaag	tgattctcct	gcctcagcct	96420
cccgagtagc	tgggactaca	ggtgcacgcc	accacgcccc	gctaattttt	tttttttttt	96480
tttttttttt	tatttgtagt	agagacagag	tttccaccatg	ttggccagga	tggtctcgat	96540
ctcctgacct	tgggatccgc	ccaccttggc	ctcccaaagt	gctagggtatt	acaggtcaga	96600
gccaccagc	ccggtgaaa	gcttttagtt	ttctaactta	tttaatttaa	tttaatttaa	96660
tttattttta	ttttattttt	tgggacagtg	tctcacttgg	ttgcccaggc	tgagtgacag	96720
tggtacaatc	atggctcact	gcacctctta	gctcctgggc	tcaggcaatc	ctcctgcctc	96780
agccttttga	gttgctggga	ctacgggcat	gtaccaccac	actcagctaa	atttttaatt	96840
ttttgtagag	atgggggtctc	actatgttgc	ctaggctggt	ctcagattcc	tcaagcattc	96900
ctcccacttg	cacctcccaa	aatgctggga	ttacaggtgt	gacaccgtgc	cagacttgaa	96960
atttttaatc	ccacacctaa	aaatataatt	ttatccacca	ttttttaaaa	gtcataacat	97020
tattttattaa	aaatttagat	ggtaaaacta	aaaattaaag	cttaataaaa	ctactgagtg	97080
atttacaagg	aagaatatta	ctggctccctt	ttgtgagcat	cccatataatt	atatatattc	97140
aggttatatt	gttacaatat	ttggttctac	tgtatacttt	tttttttgag	actgagtttc	97200
gctgttgttg	cccaggtcgg	agtgcaatgg	tgcaatctcg	gctcactgca	acctccgcct	97260
cctgggttca	agtgtattctc	ctgcctcagc	ctcctgagta	gctgggatta	caggcatgcg	97320
ccactatgcc	tggttaattt	tgtgtttttg	tagagacggg	gtttcactat	gttggtcagg	97380
ctgatctgga	actcccaacc	tcaggagatc	cgcccacctc	ggcctcccaa	agtgtcggga	97440
ttgcaggtgt	gagccaccat	gcctggctct	actatatact	ttcattcagt	tgtttctttt	97500
taatctagtg	gttttggtat	taataatttg	ataatgacct	ttagctgtta	ttgcttactt	97560
atgagttaat	atttataaag	cacttaaagt	agctggcacg	tagtaaacac	tatgtaaaga	97620
tccattaaat	aactttaaaa	aatataaaac	tgatagtggc	attttattat	agagattaaag	97680
gtaatccatt	ctctcatttc	catttatggg	atgagacgta	aacacaagta	gtttgtctctc	97740
taaaactgta	tactagaatt	ttatataccc	attatttgat	gcaactttta	taccaaagtg	97800
tattgtcagt	taccggtgaa	tatatataaa	tttaggtaag	gaaaacccaa	cttggtcgca	97860
cacagtggct	cacgcctgta	atcctaccac	tttgggaggc	cgaggcagtt	ggatcacctg	97920
aggtcttgag	ttcgagacca	gcctggccaa	cgtggcgaaa	ccccgtctcc	actaaaaata	97980
caaaaattag	ccgggcgtgg	tggcacatgc	ctgtaatccc	agctgctcgg	gaggctgagg	98040
aaggagaatt	gcttgaacct	ggggaggtgg	aggttgacgt	gagctgagat	cgtaaccactg	98100
cactccagtc	tgtgcgacag	gagcgagact	ccatctccaa	aaaaaaaaaa	aagaaaaacc	98160
caacttatct	tttacagttt	ataatagtag	aagttcaaat	aattgggttg	gaatttctgt	98220
attttttaagg	ttagtactaa	aattgttggt	tataaattgg	ggtacaatat	actttgtttt	98280
ttaagtacct	tgtgtactcta	atttaacttt	aagtccttta	ttatttttgt	ttgataagac	98340
aactttttcta	cttattttccc	cccttaactg	aaccagctac	catctgcctt	tttctgttg	98400
tatacattag	tctcttacgt	taaaatatca	tataagtttc	atatatatac	acatatcaaa	98460
ctatagactt	aaagtacaat	tacatcaaca	tcttttaaaa	ccttaatttc	tgccagggcg	98520
tggttggttca	tgccataaat	cccagcactt	tgggaggctc	aggccattgg	atcccttgag	98580
ctcaggaatt	cgagaccagc	ctggataaca	tggtgaaatc	ctcatctcta	caaaaattac	98640
aaaaagttag	ctgggcgtgg	tggcgcacac	ctatagtcct	agctacttgg	gaggatctct	98700
taagcctggg	aggcagaggt	tgacgtgagc	cgagatcatg	ccactgcact	ccagcctggg	98760
taacagttag	accctgtctc	aaaacaaaaa	tattaatttc	tataacaaat	aaattttatt	98820
tagtaatatc	atgtttaata	cccatgttac	attcattatt	cttctacact	ttggccttta	98880



ggaagaggca	gtgtgtgtgt	tacagttaaa	taggaaacaa	gtaggggttca	tatagtgtctt	98940
atgggggttt	ctttgggtggg	gggttgctgg	gtttttgggg	tttttttttt	tgagacaggg	99000
tatcactctg	tctcccaggc	tgaaatacat	tggtatgctc	atggctcact	gcagcctcaa	99060
cctccagac	tcaagtgatc	ctcccacctc	agcttcccta	gtagctgaga	ctacaggcgt	99120
gcgccaccat	gcctagctaa	attttgtaca	ttttgtagag	atgaggcttt	gccatgttgc	99180
ccaggctggt	ctctaactcc	tgggctcaag	tgatccactt	gcttcagcct	cccaaagttc	99240
gggattaaac	aggtgcgagc	cactgcagcc	agcctatatt	tttaactatat	gtttttttct	99300
tttggttaa	aatttttcag	attagtttac	aagttacaag	tgtaggtgat	atctcatgga	99360
ccaaag	atgatatgaa	gttagattgg	gtttttaaga	gtagttttta	aaatacggat	99420
ccagt	tggtggagtt	ttgtttaaag	ttctttaact	tcttattttt	ggggccaaat	99480
tatac	gttagaaatg	atttttaaca	caggctcattt	atgccaaact	gcattttgcc	99540
ttttt	tttttttt	tttttttgag	agggagtctc	actcattgcc	caagctggag	99600
tgga	atctcggc	tcacttacaa	cctccacctc	ccaggttcaa	gcgattctcc	99660
ccagcc	tcttgagtag	ctgggattac	aggcacctgc	ctggctaatt	tttgattttt	99720
ccagac	gggggtttcac	catgttggcc	aggctggtct	cgaactgctg	acctcagggtg	99780
ccagc	cttgccctcc	caaagtgcgg	ggattacagg	tgtgagccac	cgtgcttgcc	99840
cccttt	attaattttt	ctattgcctg	gactctgtga	acctatccat	tttgccctttt	99900
ctactt	aggtgtaaat	atagatatct	attaactcag	cattgtttta	attatatttt	99960
aggcaa	tttaaaagat	cagaaaaata	gaccaaatta	atataaaaat	gcatacttta	100020
gggcaa	agtggctcac	gtctgtaatc	ccagcacttt	gggaggccaa	ggtgggcgga	100080
ctgagg	ttaggagttc	aggaccaacc	tggccatcat	ggcgaaaacc	tgtctctact	100140
ctacaa	aaattagctg	ggcatggagg	catgtgcctg	taatcccagc	tactcgggag	100200
ccagac	aagaattgct	tgaaccttgg	agggggagggt	tgcatatctg	agtggtgaaa	100260
ctgtgattct	ttttttctct	ttgtctgtat	ttttgaactt	ttctataaat	gatttgtgtt	100320
tgtttttata	ttggaaaaat	attatgcttt	caaagtgtta	tacctatgaa	actaaacaca	100380
agtaataaat	atattagtat	agcattttatt	aaggtttctt	gtgtagcaga	tcaacataga	100440
aaatatattt	aaatggctga	cataattttc	taagaatata	tacacgtata	ttttttataa	100500
cattaagaaa	cagcagcatc	attactttta	tccatcattt	cgttaaccac	catataccctg	100560
ttgatcattt	aatgtgtcat	gtgcttttta	aaaatctaga	tgagaaatat	tcgattatct	100620
gacttacttg	aatccttgaa	aaaaataaaa	cgcagcgtca	gccctcaaac	tttagaagcg	100680
tacatacgtt	ggaacaagga	ctttggagat	accactgttt	aaggaaatac	ctttgtaaac	100740
ctgcagaaca	ttttacttaa	aagaggaaac	acaagatctt	caatgaacgt	catcggctac	100800
agaaacagcc	taagtttaca	ggacttttta	gagtcttaca	tatttggtgca	ccaaacttga	100860
agatgaacca	gaaaacagac	ttaaacaaaa	tatacaatgc	aaatgtaatt	ttttgttggt	100920
taaggccctg	ccttgatggt	cacagttatc	ccaatggaca	ctaagttaga	gcacaacaaa	100980
acctgattct	ggtcttcttt	accaatataa	tcataatgta	aataataatt	tgtatatgtg	101040
gttgacagatg	aaagtattcc	aggaacagtg	aatggtagaa	gacacaagaa	catttgtttg	101100
tttgctcttg	gatgtttttt	cttaaaatag	taatttctcc	tacttttctt	ttctactgtt	101160
gtcttaacta	cagggtgattg	gaatgccaaa	cactcttaag	tttattttct	ttttctgttt	101220
tataaattca	gtgtgccaaa	tgaaactttt	ttcctaagta	actgtaatag	gaaaaagtgt	101280
attttgagag	tttcttcttc	ataaatctac	agacattaaa	caattgttgt	gttcttttta	101340
ccttttattt	ttctattacc	ttgctaccaa	acagtttaga	tagcaatata	atagcaaaaa	101400
agcaaatatg	gtaaaataga	gaaggtttga	aggtttgagt	tactctgtca	tataacatgt	101460
agatcagtc	tcatgtgacc	tgacgtattt	ttttttctaa	tgatattgtc	agaaatctgt	101520
tgtagactgt	taacttcttc	ctgatggaat	ttattttctg	caagaattat	tctgatattt	101580
aagagagcca	attttaactg	ctgtgaaaat	gtttccagtg	caagagaagg	gaaatactag	101640
gaactaagac	attttctaatt	tattgcttat	tactttctta	atttttacagg	ataattataa	101700
gcaagtggaa	ctaccatctt	ttattcttaa	taattattaa	tcccttcaat	gaaactttta	101760
aaaaactgaa	tttttatata	tggcatacat	ttttctagtt	ccttctgctt	gctttattaa	101820
ctcaaaagt	ctagtcttag	tctgttgatc	tgctttttgt	tctcccaaaa	tgtacagtaa	101880
ttccatttgt	ttgtataaat	atgcctggat	tttcattata	aaaatgtcat	tgtagggagt	101940
agagactcat	atcatggcct	tttaaatatt	gtaataaagg	caaataagata	tttgccctta	102000
gtttactggt	taaaagtgtg	tttacagaac	ttttctctgg	tgcttaaatg	atgctatgta	102060
aaatgtcatg	agtggaaaga	atattttag	tagtaacaag	aatttttcat	ttaggaaaga	102120
tttcttaggt	tttgaaagaa	tacattaaaa	taaaaaactt	gccctacta	ggtaagaact	102180
ttataatgaa	gacatacatt	cttcttaatt	ttactcttgc	tcttggttaa	gatttgtttg	102240
aatatagaag	atgcatgatt	tctgggtttt	tttttttttt	tgagacagag	tttcgctctt	102300
gttgccagg	ctggagtga	atggcgcaat	ctcgactcac	cacaacctcc	gcctcccagg	102360
ttcaagcaat	tctcctgcct	cagcctcccg	agtagctggg	attacaggca	tgcgccacta	102420



ccccagctaa	ttttgtat	ttagtagaga	tgggggttct	ccatgttggt	caggctgggc	102480
ttgaatcct	gacctcaggt	gatccgcctg	cctcggcctc	ccaaagtgc	gggattacag	102540
gcataagcca	ctgcgccag	ccagaagatg	catgatttct	taggatcata	tgctgtttgt	102600
agccataagg	taaatcatgt	ctcttccaat	catgactttg	gaactccctg	aataataaaa	102660
atgagagttg	agataaatag	gggaaaaaaa	atTTTTTTca	agccagagct	atgcataatg	102720
taggtgatgg	gtagtatccc	tttaaggtct	caaacattac	aacatcaatt	atgaaatact	102780
gataacgaaa	ggtagtaatg	aaatatatat	gatgaaaaga	attgagaagt	tctaaattaa	102840
gacatttcag	ttaaagctcat	aaaatttcat	tgttttccatt	taaaagatta	acgttattga	102900
tacttgata	actggcta	catattaaag	gactatgtgg	ttccagctca	acttttaata	102960
tattgtctcc	tttaaaacta	tcatggttat	aattctattg	ggaaagactt	ttagataaca	103020
aagatttcaa	atgttaaaag	agataaaagt	cagggttaata	ctatcttaaa	cactgagtc	103080
gaaaatcatt	actgtataga	agttgcttct	ctgatcaagt	ctgaacttca	gctagtgc	103140
gagaactatt	ttctatgact	taactcta	caagttttat	tttaagctgt	ttctttgata	103200
gaagggccat	gaaaatagag	taatgatata	gtaggagata	agggattgg	ttggctcttt	103260
tcaataaaga	tagaagttgc	tgaagtttct	tgaattaata	atgacttaga	ttgtgacctt	103320
ttagattcgg	tggtgagctc	tggtgtgtat	tacttcttaa	aagataatgc	ttaaacatta	103380
agcattagtg	tgctcttcat	gttaatatgg	cagagttttg	taaactaaat	taaaacttac	103440
tgatatattg	gactttgagc	caagggaaa	aatgagtact	atctttccag	atatcttaag	103500
ggtaaaagct	tattctaaga	cagtctgtcc	attgagaata	ttagatttct	gacttgcaaa	103560
tatgtttgta	ctccagaaga	attagaggaa	aagcagatac	tagaattcta	atttaattac	103620
atatacagcc	gtctttgttt	atagtgtaga	attctttata	ttttgtacaa	aaactaattc	103680
ttttggtaaa	atgaaccatt	tacagttcgg	ttttggactc	tgagtcaaa	gattttctct	103740
taaatgcttg	tctcaatttt	agtctgggtc	tttgtacttt	tcttcagaag	aaatgaatta	103800
aagggtacag	ttgcataaag	tgggttttta	tcctaagtga	ttggaaataa	atgataaact	103860
ttattttgtc	tttacttttt	tactttaaaa	ctttttgata	ttttaggggt	tggagtctga	103920
taatgaagga	gttgtgtgta	ttggactctt	agtaacaatt	ataaacgctt	aacaaaatat	103980
agaaagcaat	gattgggtgg	ctctgcagag	caatcaaac	aaggtagaaa	ctgcaaagtc	104040
ctataatgga	agagattcag	gctgggtgtg	gtggctcatg	cctgtttggg	aggccagtat	104100
gggagaatcg	cttacactca	ggaattcaag	accagcctgg	ggaatatagg	gagaccctgt	104160
ttctgccaaa	aaaaaaaaaa	aaaaaaaaac	taaactaaaa	attagccagg	tttgggtggc	104220
tacacctgta	gtcccggtta	ctcaagaggc	tgagggttaga	ggatcgctta	agcccgggag	104280
gaggaggttg	cagtgaagcca	agatcacacc	accacactcc	agcccagggt	acagagaacc	104340
tgctcctaaa	aataaaaagt	aaacgagatt	caccttaacc	agcttttacc	catagggcaa	104400
tttccagttt	gtgcagttca	cttggatata	gaattcaggt	agaaatgtgac	agttttgaca	104460
ggctgaggaa	tcatttgga	cttctagaac	acctgaaaat	tagaagagaa	attttgga	104520
ggagagtgcc	tcagaaagta	agctctaaaa	tctgcctaca	tattcctttc	aatccttg	104580
tggattccaa	aattgtgtctg	gtgcagtg	attattttaa	ggaaccagg	agaaagcaag	104640
ttaaagtcta	aaaaactcag	cagcgattgc	agctgcccaa	ggacagagag	tttgaaattc	104700
aagttccact	aagaaggact	gagtaaatac	ttgggtgttc	ccattgaacc	cccaaatgc	104760
cacgccttaa	ggaagaatga	cagcatccta	gaactaaagg	ctgtgcttca	gcactaagga	104820
caaatggaa	ataaacttac	tgtaatgaag	cttaaaatca	aatctcacag	catcaaggtt	104880
atccattagt	aattttaa	cagaacaaaa	cacagcattc	ttgagaagaa	aacaattcag	104940
tcatacagc	atatatccaa	agctcagtac	ataataaaca	gttactaagc	atgcaaagaa	105000
aaattatgtg	atccttttga	ctcaacttcc	ctgatggaaa	aaaaaaagt	gtccatacca	105060
tactcaacag	aaaatataca	gaagcagatg	cacaagtga	ctctatgaga	gtgattttaa	105120
aatagcgtaa	aaatatgtta	aaagaattta	caagaagaga	taatgggtga	atagatagg	105180
agtttcagca	gaaaaataag	aattgaaaac	ccaacagagg	ccaggcaccg	tgctcacgc	105240
ccgtaatcac	agcacttttg	gaggccaagg	caggtggatc	acctaaaggt	aggagttaa	105300
gaccagcctg	accaacatgg	tgaaccgg	tctctactaa	aaatacaaaa	ctagccgggc	105360
atgggtgatg	catgcccata	atcctagcta	cttggggaggc	tgaagcagga	gaattgcttg	105420
aacctgggag	gtagaggttg	cagtgaacca	agatcatgcc	attgcactcc	agcctgggca	105480
agaagagcaa	aactccatct	caaaaaataa	taataataat	tggtggggcg	cgggtggctc	105540
cgctgtaat	cccagcactt	tgggagccg	agacaggcag	atcacaaggt	caggagattg	105600
agaccatcct	ggctaacatg	gtgaaacccc	gtctctacta	aaaatacaaaa	aaattagcct	105660
ggcatgggtg	cgggcacctg	tagtcccagc	tactcgggag	gctgaagcag	gagaatggca	105720
tgacccagg	aggcggaact	tgcaatgagc	ccagatcgcg	ccactgcact	ccagcctggg	105780
caaccgtctc	aaaaataata	ataataataa	ttagaaaata	ataaaaaata	acccactgga	105840
aagtcttgta	atgaaaacta	cagtatctga	aattaaaaat	ttaaatggat	ggacttaaca	105900
gtaaactagg	cacaacagga	aagactgaaa	aagacaagta	agtatcaaac	caaagcagag	105960



agagaaagaa	aaaagaaata	gcagagggtt	gagacctttg	gaacattatc	agttcctgta	106020
atttgaagta	aagaaggaaa	ggaagagaat	ggggcaagag	aaatctttga	agcaatgatg	106080
cgtgaaat	tccccaaggt	gttgaaagac	atcaacgtac	agatctaaga	agttcactga	106140
atccccaagc	agaataaata	caacaaagac	tacatctggg	tacatcacat	tgccaaaaaa	106200
aaaaaaaaaa	aaaatcttaa	aagcaatcaa	ggtggggagt	gggaaggagt	agcattacat	106260
tcagataaac	aagagtgatg	gctaacctca	tcagaaatga	tggaattaag	aaaatgatgg	106320
aatgacatct	ttaagggtgca	aagatgaaac	aaagggttaac	ttaaaaactct	gtatccagtt	106380
aaaatatcct	tcaagatggt	aaggcaatgg	ccaggctcag	tggttatatgc	ctgtaatccc	106440
aacacttttg	ggggctgaga	caggagaatc	ccttgaggcc	atgagtgacc	agcctgggca	106500
acctagttag	aaccccatca	aaaaataaaa	ttagctgggc	atgggctgtg	gtcgtgctac	106560
tgcactccag	cctgggggac	agagtcccta	tctcttgaaa	aacttaagac	aaaatctttt	106620
ccagatgaaa	acagaatttg	ttgctggcag	actgacttta	tacggaatgc	tttaagaaat	106680
tcttcaggct	gaaggaaaaat	ataaatctgg	agtgaaaaac	gtagtatctg	aaattaaaaat	106740
tttcttggga	tggaacttatc	agtaaattgga	tttcccaatg	ggaacctgta	tctccaggat	106800
ggaattaaga	atthttgaaat	gataaacatg	ttggtaaata	agaaagacat	ttttcttttt	106860
tatthtttcta	aatatcattg	ctgttttaag	caaacaaaaa	tagtattgca	gtgtttataa	106920
tatatgtata	agtaaaatat	aggacagaaa	tagcatgaag	ggtcagagag	ggcaataaat	106980
gtaattatat	tgagataata	ttactatatt	gtacacaaag	tgagatagca	ttagtgtag	107040
gtagattgtg	ataagaatgc	atthttgtaat	ctctcacccc	atcactgggg	tgcatgatag	107100
ctaaaaagcc	attagagaag	ataaaatggg	tcactaagaa	ttatgtgatt	agtc aaatga	107160
aatcaaggaa	agaaaaacag	agcaaagaag	tgtaagataa	aacagtggga	tggaagagct	107220
gaaattaata	atgtcactaa	attgcttacc	ttcctccttc	ctgaactgtt	gtggattttt	107280
tttttttcag	gggggttagg	ttttttttta	gagacagagt	ctctattgcc	ccaggctgga	107340
cagcagtggc	gtaatcagtc	tcaaaatccc	tgcaagtctca	aactcctggg	ctcaagtgat	107400
cctcctgcct	cggcctccca	aatgctggg	attacaggca	ggcagagcca	ccacaccctg	107460
ccagaactat	tgthtttaag	tcattaggtg	gttaagccct	cgagactatg	tgcaagtttg	107520
atgattcggt	aggagcactc	acaggactca	gcatatagtc	ttattcccaa	caatgattta	107580
ttagagcaaa	aggatatgaa	gcaaattcag	caaagggaag	aaaggaatta	agtgaagtcc	107640
agaggaaacc	aagcaccagc	ttctaagggt	tctctcctaa	tgaagtcaca	caggatgtca	107700
ctccagcaac	aagtgcagac	atgtgaaagt	gttaagccct	actgttagac	atagtaagct	107760
tggttagggaa	tagagggaat	ctcccctcaa	attcaagttc	ccagatgcct	ggcaagggcc	107820
aaccttgcca	gcaggacttt	ctggggatag	cagtcccagg	cacgtgtttg	cacagtgggt	107880
cagatcaatg	tgcttacatt	gggttagagg	gacctatgga	agtccaaatt	tggtgtgcag	107940
agacctataa	gggtgaagac	agtgtctaca	atgatggaca	gctaggtatg	aggtgtcagg	108000
gacagattga	ggcagttatt	cacatggtgg	ataggggcaa	cctggaatga	ggaataagct	108060
caagcacaga	aaggggtgtc	catatgggg	aaggggtgtca	gcagagatgg	tagattgggt	108120
gcacaccaaa	gacttgatgg	aataaaagtga	atatactaac	cacagagaag	gtaattataa	108180
atacagaag	gggggaaact	agaatgaaac	ctggagttca	gcttgaattg	ggtttaagaa	108240
agtgaattca	tggttttaaaa	tctataaaaa	tagatgaaat	atcggctggg	cacagtggct	108300
cacgcataata	atcccagcac	tttgggaggc	cgaggcgggc	agatcgcttg	agatcaggag	108360
tttgagacca	gcctggccaa	cactgcgaaa	ccccatctct	agtaaaaaata	caaaaattag	108420
ccagggtgtg	tggtacacac	ctgtagtccc	agctacttgg	gaggctgagg	cacgagaatc	108480
acttgaacct	gggaggcaga	ggttgcaagt	ggctgagatc	gcaccactgc	actccagcct	108540
gggcaacaga	gcgagatcct	gcctcaaata	aaaatagagg	acotataagc	atataaatat	108600
acacgtgtgt	atgtgtccaa	atatgtatat	tccctagtct	gtccaccaag	gtggccttgg	108660
agcagttatg	ctccaataat	aatgagcaca	taaagtaccc	atatcttgcc	ttccaaattc	108720
ttcactgtct	tagtctgctt	gggctgcatt	acaaaatacc	atagactggg	cagcttaaat	108780
aacagaaatt	tatcttcaca	gttctagaag	cctggaagtc	caagattaaa	gtaccagcca	108840
gthttggttt	tagtgagggc	tttcttctcg	gcttgcaagt	ggccaccttc	tcaccgtgtc	108900
cttgatggc	agacagcaca	agctctctgg	tgtccctttt	taaaagggca	ttaatcccat	108960
catgacagtc	ccatcctcat	tatctcatct	aaccctaggt	acttcccaaa	ggctgaatca	109020
ccaaagacca	tcacattgct	ggtgaaggct	tcaacatatg	aatttgaggg	acacgaatat	109080
tcagtccata	acatcaacta	aaggaaccaa	gactctttga	taaaatggct	aaattcaggg	109140
ctggggcaga	gaaaatacat	gagtgtggaa	cttcttgtgc	cagagagaaa	aagtgcctaa	109200
agattgatga	ggatgaatca	ttgaaatgac	acacagatta	aaagggttcc	cactggacaa	109260
atthtgagcat	caaaaataagt	aatagtagta	attaattata	acccatcaga	agaaaataac	109320
catgagctca	tgtgaatata	tgaatacaaa	cataaacaaa	ttacaagcat	aatgaggaat	109380
gtgatattta	tatgggttaa	aggtacctct	ccaggccggg	tgcaagtaact	ctcacctgta	109440
atcccagcac	tttgggaggg	caaggcaggt	agatcacctg	aggtcaggca	tttgagacca	109500



gcctgcacaa	catggtgaaa	ccctgactct	actaaaaata	cataacgcga	gccgggctg	109560
gtggcacgtg	tctataatct	gccactgatt	aggtgtgtga	ttttcccaag	caggggataa	109620
tagtagtacc	tatgtcaaa	gctgttatga	ggattaaatg	agctaacaca	taatcgtgct	109680
tttttttttt	tttttttttt	ttgagacaga	gtcttgact	gtcgcctggg	ctggagtgc	109740
atggcacgat	ctcgccccc	tgcaacctct	gcctcccagg	ttcaagtgat	tctcctgcct	109800
cagcctcctg	agtagctggg	attacaggct	cctgccacca	cacctggcta	ttttcaatag	109860
agacgggggt	tcactatgtt	ggccaggcta	gtctcaaaaa	cctgacctcg	tgatccaccc	109920
gctttggcct	cccaaagtgc	tgggattaca	ggcatgagcc	actgcacccg	gctttttttt	109980
tttttttttt	agatggaatc					110000

<210> 2  
 <211> 3263  
 <212> DNA  
 <213> Homo sapiens

<400> 2						
gctcctgaga	ccggcgggca	cacgggggtc	tgtggccccc	gccgtagcag	tggctgccgc	60
cgctcgcttg	ttcccgcggg	tctgcgggag	gcgggttatg	gcggcggcgg	cagtgcagagc	120
tgtgaatgaa	ttctccgggt	ggacgagggg	agaagaaagg	ctccggcggc	gccagcaacc	180
cggtgcctcc	caggcctccg	cccccttgcc	tggccccgc	ccctcccgc	gccgggcccgg	240
ccccccggcc	cgagtcgccc	cataagcgga	acctgtacta	ttctcctac	ccgctgtttg	300
taggcttcgc	gctgctgcgt	ttggtcgcct	tccacctggg	gctcctcttc	gtgtggctct	360
gccagcgctt	ctcccgcggc	ctcatggcag	ccaagaggag	ctccggggcc	gcgccagcac	420
ctgcctcggc	ctcggccccc	gcgcgggtgc	cgggcggcga	ggccgagcgc	gtccgagtct	480
tccacaaaca	ggccttcgag	tacatctcca	ttgccctgcg	catcgatgag	gatgagaaag	540
caggacagaa	ggagcaagct	gtggaatggt	ataagaaagg	tattgaagaa	ctggaaaaag	600
gaatagctgt	tatagttaca	ggacaagggt	aacagtgtga	aagagctaga	cgccttcaag	660
ctaaaatgat	gactaatttg	gttatggcca	aggaccgctt	acaacttcta	gagaagatgc	720
aaccagtttt	gccattttcc	aagtcacaaa	cggacgtcta	taatgcagat	actaacttgg	780
catgccgcaa	tggacatctc	cagtcagaaa	gtggagctgt	tccaaaaaga	aaagaccctt	840
taacacacac	tagtaattca	ctgcctcggt	caaaaacagt	tatgaaaact	ggatctgcag	900
gcctttcagg	ccaccataga	gcacctaagt	acagtggttt	atccatgggt	tctggagtga	960
aacagggatc	tggctcgtct	cctaccatc	ataagggtac	tccgaaaaca	aataggacaa	1020
ataaaccttc	tacccctaca	actgctactc	gtaagaaaaa	agacttgaag	aatttttagga	1080
atgtggacag	caaccttgct	aaccttataa	tgaatgaaat	tgtggacaat	ggaacagctg	1140
ttaaatttga	tgatatagct	ggtcaagact	tggcaaaaaca	agcattgcaa	gaaattgtta	1200
ttcttccttc	tctgaggcct	gagttgttca	cagggcttag	agctcctgcc	agagggctgt	1260
tactctttgg	tccacctggg	aatgggaaga	caatgctggc	taaagcagta	gctgcagaat	1320
cgaatgcaac	cttctttaat	ataagtgtcg	caagtttaac	ttcaaaaatac	gtggggagaag	1380
gagagaaaatt	ggtgagggct	ctttttgtcg	tggctcgaga	acttcaacct	tctataattt	1440
ttatagatga	agttgatagc	ctttttgtgtg	aaagaagaga	aggggagcac	gatgctagta	1500
gacgcctaaa	aactgaattt	ctaatagaat	ttgatgggtg	acagtctgct	ggagctgaca	1560
gagtacttgt	aatgggtgca	actaataggc	cacaagagct	tgatgaggct	gttctcaggc	1620
gtttcatcaa	acgggtatat	gtgtctttac	caaattgagga	gacaagacta	cttttgctta	1680
aaaaatctgtt	atgtaaacaa	ggaagtccat	tgacccaaaa	agaactagca	caacttgcta	1740
gaatgactga	tggatactca	ggaagtgacc	taacagcttt	ggcaaaaagat	gcagcactgg	1800
gtcctatccg	agaactaaaa	ccagaacagg	tgaagaatat	gtctgccagt	gagatgagaa	1860
atattcgatt	atctgacttc	actgaatcct	tgaaaaaaat	aaaacgcagc	gtcagccctc	1920
aaactttaga	agcgtacata	cgttggaaca	aggactttgg	agataccact	gttttaaggaa	1980
atacctttgt	aaacctgcag	aacattttac	ttaaaagagg	aaacacaaga	tcttcaatga	2040
acgtcatcgg	ctacagaaac	agcctaagtt	tacaggactt	tttagagtct	tacatatattg	2100
tgcaccaaac	ttgaagatga	accagaaaa	agacttaaac	aaaatatata	atgcaaatgt	2160
aattttttgt	tgtttaaggc	cttgccctga	tggtcacagt	tatcccaatg	gacactaagt	2220
tagagcacaa	caaaacctga	ttctgggtctt	ctttaccaat	ataatcataa	tgtaataaat	2280
aatttgtata	ttgtgttgca	gatgaaagta	ttccagggaac	agtgaatggt	agaagacaca	2340
agaacatttg	tttgtttgtc	ttctgatgtt	ttttcttaaa	atagtaattt	ctcctacttt	2400
tcttttctac	tgttgtctta	actacagggtg	attggaatgc	caaacactct	taagtttatt	2460
ttcttttttc	gttttataaa	ttcagtggtgc	caaatgaaac	ttttttccta	agtaactgta	2520
ataggaaaaa	gtttattttg	agagtttctt	cttcataaat	ctacagacat	taaacaattg	2580



ttgtgttctt	tttacctttt	atTTTTctat	taccttgcta	ccaaacagtt	tagatagcaa	2640
tataatagca	aaaaagcaaa	tatggtaaaa	tagagaaggt	ttgaagggtt	gagttactct	2700
gtcatataac	atgtagatca	gtcttcatgt	gacctgcagt	attttttttt	ctaattgtatt	2760
tgtcagaaat	ctgttgtaga	ctgtttaaactt	cttcctgatg	gaattttattt	tctgcaagaa	2820
ttattctgat	atttaagaga	gccaattttta	actgctgtga	aaatgtttcc	agtgaagag	2880
aagggaata	ctaggaacta	agacattttct	aattttattgc	ttattacttt	cttaattttta	2940
caggataatt	ataagcaagt	ggaactacca	tctttttattc	ttataataatta	ttaatccctt	3000
caatgaaact	ttaaaaaaac	tgaatttttta	tacatggcat	acattttttct	agttcccttct	3060
gcttgcttta	ttaaactcaaa	agttctagtt	ctagtctgtt	gatctgcctt	ttgttctccc	3120
aaaatgtaca	gtaattccat	ttgtttgtat	aaatatgcct	ggattttcat	tataaaaaatg	3180
tcattgtagg	gagtagagac	tcatatcatg	gcctttttaa	tattgtaata	aaggcaata	3240
gatatttgcc	cttagtttac	tgg				3263

&lt;210&gt; 3

&lt;211&gt; 616

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 3

Met	Asn	Ser	Pro	Gly	Gly	Arg	Gly	Lys	Lys	Lys	Gly	Ser	Gly	Gly	Ala	
1				5				10					15			
Ser	Asn	Pro	Val	Pro	Pro	Arg	Pro	Pro	Pro	Pro	Cys	Leu	Ala	Pro	Ala	
		20					25					30				
Pro	Pro	Ala	Ala	Gly	Pro	Ala	Pro	Pro	Pro	Glu	Ser	Pro	His	Lys	Arg	
		35					40					45				
Asn	Leu	Tyr	Tyr	Phe	Ser	Tyr	Pro	Leu	Phe	Val	Gly	Phe	Ala	Leu	Leu	
	50					55					60					
Arg	Leu	Val	Ala	Phe	His	Leu	Gly	Leu	Leu	Phe	Val	Trp	Leu	Cys	Gln	
	65				70					75					80	
Arg	Phe	Ser	Arg	Ala	Leu	Met	Ala	Ala	Lys	Arg	Ser	Ser	Gly	Ala	Ala	
			85						90					95		
Pro	Ala	Pro	Ala	Ser	Ala	Ser	Ala	Pro	Ala	Pro	Val	Pro	Gly	Gly	Glu	
			100					105					110			
Ala	Glu	Arg	Val	Arg	Val	Phe	His	Lys	Gln	Ala	Phe	Glu	Tyr	Ile	Ser	
		115					120					125				
Ile	Ala	Leu	Arg	Ile	Asp	Glu	Asp	Glu	Lys	Ala	Gly	Gln	Lys	Glu	Gln	
	130				135						140					
Ala	Val	Glu	Trp	Tyr	Lys	Lys	Gly	Ile	Glu	Glu	Leu	Glu	Lys	Gly	Ile	
	145				150				155						160	
Ala	Val	Ile	Val	Thr	Gly	Gln	Gly	Glu	Gln	Cys	Glu	Arg	Ala	Arg	Arg	
			165					170						175		
Leu	Gln	Ala	Lys	Met	Met	Thr	Asn	Leu	Val	Met	Ala	Lys	Asp	Arg	Leu	
		180						185					190			
Gln	Leu	Leu	Glu	Lys	Met	Gln	Pro	Val	Leu	Pro	Phe	Ser	Lys	Ser	Gln	
		195					200					205				
Thr	Asp	Val	Tyr	Asn	Asp	Ser	Thr	Asn	Leu	Ala	Cys	Arg	Asn	Gly	His	
	210				215						220					
Leu	Gln	Ser	Glu	Ser	Gly	Ala	Val	Pro	Lys	Arg	Lys	Asp	Pro	Leu	Thr	
	225				230					235					240	
His	Thr	Ser	Asn	Ser	Leu	Pro	Arg	Ser	Lys	Thr	Val	Met	Lys	Thr	Gly	
			245						250					255		
Ser	Ala	Gly	Leu	Ser	Gly	His	His	Arg	Ala	Pro	Ser	Tyr	Ser	Gly	Leu	
			260					265						270		
Ser	Met	Val	Ser	Gly	Val	Lys	Gln	Gly	Ser	Gly	Pro	Ala	Pro	Thr	Thr	
		275					280						285			
His	Lys	Gly	Thr	Pro	Lys	Thr	Asn	Arg	Thr	Asn	Lys	Pro	Ser	Thr	Pro	
	290					295					300					



Thr Thr Ala Thr Arg Lys Lys Lys Asp Leu Lys Asn Phe Arg Asn Val  
 305 310 315 320  
 Asp Ser Asn Leu Ala Asn Leu Ile Met Asn Glu Ile Val Asp Asn Gly  
 325 330 335  
 Thr Ala Val Lys Phe Asp Asp Ile Ala Gly Gln Asp Leu Ala Lys Gln  
 340 345 350  
 Ala Leu Gln Glu Ile Val Ile Leu Pro Ser Leu Arg Pro Glu Leu Phe  
 355 360 365  
 Thr Gly Leu Arg Ala Pro Ala Arg Gly Leu Leu Leu Phe Gly Pro Pro  
 370 375 380  
 Gly Asn Gly Lys Thr Met Leu Ala Lys Ala Val Ala Ala Glu Ser Asn  
 385 390 395 400  
 Ala Thr Phe Phe Asn Ile Ser Ala Ala Ser Leu Thr Ser Lys Tyr Val  
 405 410 415  
 Gly Glu Gly Glu Lys Leu Val Arg Ala Leu Phe Ala Val Ala Arg Glu  
 420 425 430  
 Leu Gln Pro Ser Ile Ile Phe Ile Asp Glu Val Asp Ser Leu Leu Cys  
 435 440 445  
 Glu Arg Arg Glu Gly Glu His Asp Ala Ser Arg Arg Leu Lys Thr Glu  
 450 455 460  
 Phe Leu Ile Glu Phe Asp Gly Val Gln Ser Ala Gly Asp Asp Arg Val  
 465 470 475 480  
 Leu Val Met Gly Ala Thr Asn Arg Pro Gln Glu Leu Asp Glu Ala Val  
 485 490 495  
 Leu Arg Arg Phe Ile Lys Arg Val Tyr Val Ser Leu Pro Asn Glu Glu  
 500 505 510  
 Thr Arg Leu Leu Leu Leu Lys Asn Leu Leu Cys Lys Gln Gly Ser Pro  
 515 520 525  
 Leu Thr Gln Lys Glu Leu Ala Gln Leu Ala Arg Met Thr Asp Gly Tyr  
 530 535 540  
 Ser Gly Ser Asp Leu Thr Ala Leu Ala Lys Asp Ala Ala Leu Gly Pro  
 545 550 555 560  
 Ile Arg Glu Leu Lys Pro Glu Gln Val Lys Asn Met Ser Ala Ser Glu  
 565 570 575  
 Met Arg Asn Ile Arg Leu Ser Asp Phe Thr Glu Ser Leu Lys Lys Ile  
 580 585 590  
 Lys Arg Ser Val Ser Pro Gln Thr Leu Glu Ala Tyr Ile Arg Trp Asn  
 595 600 605  
 Lys Asp Phe Gly Asp Thr Thr Val  
 610 615

&lt;210&gt; 4

&lt;211&gt; 23

&lt;212&gt; ADN

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Amorce

&lt;400&gt; 4

cggagctcct cttggctgcc atg

23

&lt;210&gt; 5

&lt;211&gt; 26

&lt;212&gt; ADN

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Amorce



<400> 5  
agaagcgctg gcagagccac acgaag

26

<210> 6  
<211> 27  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 6  
aaggcgacca aacgcagcag cgcgaag

27

<210> 7  
<211> 26  
<212> ADN  
<213> Artificial Sequence

<220>  
<223> Amorce

<400> 7  
aggagcaagc tgtggaatgg tataag

26

<210> 8  
<211> 27  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 8  
tggttatggc caaggaccgc ttacaac

27

<210> 9  
<211> 26  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 9  
caaacggacg tctataatga cagtac

26

<210> 10  
<211> 25  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 10  
ttaggaatgt ggacagcaac cttgc

25

<210> 11  
<211> 25  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce



<400> 11  
cttctctgag gcctgagttg ttcac

25

<210> 12  
<211> 27  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

12  
g actgatggat actcagg

27

icial Sequence

Amorce

> 13  
cagca ctgggtccta tccg

24

> 14  
<211> 26  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 14  
atgaacgtca tcggctacag aaacag

26

<210> 15  
<211> 19  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 15  
tagcagtggc tgccgccgt

19

<210> 16  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 16  
aagcggctct tggccataac

20

<210> 17  
<211> 19  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce



<400> 17  
ggcggcagtg agagctgtg 19

<210> 18  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 18  
ctagctcttt cacactgttc 20

<210> 19  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 19  
aacaggcctt cgagtacatc 20

<210> 20  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 20  
ctgtgaacaa ctcaggcctc 20

<210> 21  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 21  
atgagaaagc aggacagaag 20

<210> 22  
<211> 18  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 22  
tgccaagtct tgaccagc 18

<210> 23  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce



<400> 23  
ctacaactgc tactcgtaag 20

<210> 24  
<211> 19  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 24  
cagtgtgca tcttttgcc 19

<210> 25  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 25  
taggaatgtg gacagcaacc 20

<210> 26  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 26  
aaagctgtta ggtcacttcc 20

<210> 27  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 27  
tggagatgac agagtacttg 20

<210> 28  
<211> 19  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 28  
ctggaatact ttcactgc 19

<210> 29  
<211> 19  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce



<400> 29  
atgaggctgt tctcaggcg 19

<210> 30  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 30  
gtgagccgaa ctgcacattg 20

<210> 31  
<211> 21  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 31  
caaagtcgac agctacagtg c 21

<210> 32  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 32  
ggaactgtag ttgagtggga 20

<210> 33  
<211> 19  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 33  
agatgaggct ccgacctac 19

<210> 34  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 34  
aatgccacac ttgtaatctc 20

<210> 35  
<211> 22  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce



<400> 35  
tgtgaatata tcataatttg gg 22

<210> 36  
<211> 19  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 36  
tacagcagtt ctcatgatg 19

<210> 37  
<211> 21  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 37  
gaccaaattg gtgcatgcat g 21

<210> 38  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 38  
acatttccaa tacatcccac 20

<210> 39  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 39  
atttgtcatt tcacatgcac 20

<210> 40  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 40  
ttagaatgac tatacctgac 20

<210> 41  
<211> 18  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce



<400> 41  
tcagggttaag taagactc 18

<210> 42  
<211> 19  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 42  
ttcctatcta cctagtgac 19

<210> 43  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 43  
ttttatagca agttgccctg 20

<210> 44  
<211> 19  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 44  
cctatgaaga tcctggtac 19

<210> 45  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 45  
tgtcatgatt ctaacaagg 20

<210> 46  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 46  
tctatttcac tcctgacatg 20

<210> 47  
<211> 19  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce



<400> 47  
gtcatagggc ttaggcttc 19

<210> 48  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 48  
atcatactac ccacttttcc 20

<210> 49  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 49  
tgtttgggaa gatgctactg 20

<210> 50  
<211> 21  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 50  
ctactgaaga taacgtacat g 21

<210> 51  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 51  
cattgattgc catgtattgg 20

<210> 52  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 52  
agaaggccag aaatactcag 20

<210> 53  
<211> 22  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce



<400> 53  
gtacttaaatt cggtaaaatat gg

22

<210> 54  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 54  
aaagtctt aggaatgcag

20

## Sequence

55  
acttaacc aggctgtatg

20

<210> 56  
<211> 19  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 56  
ctcagatgac tcacatagc

19

<210> 57  
<211> 22  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 57  
ctttactaga ctaattctcc tg

22

<210> 58  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 58  
cagattcaag aagacagatc

20

<210> 59  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce



.

.

.

.

<400> 59  
gcaataattc accacacttg 20

<210> 60  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 60  
ggtagttctt gtttctgctc 20

<210> 61  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 61  
caagtgtggt gaattattgc 20

<210> 62  
<211> 19  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 62  
gagctgaaaa gtattcagc 19

<210> 63  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 63  
tgcaaaggac atagccagtg 20

<210> 64  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 64  
agcctctgga gatagtatgc 20

<210> 65  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce



<400> 65  
ctagaacagg ggtcacagtc

20

<210> 66  
<211> 18  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 66  
ttggacttct taaacttc

18

<210> 67  
<211> 21  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 67  
gcagtatgca agaaattgaa c

21

<210> 68  
<211> 19  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 68  
ggcctgtaat tttcttctg

19

<210> 69  
<211> 21  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 69  
gtactgaata gatacatgta g

21

<210> 70  
<211> 19  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce

<400> 70  
gtgtagcaga tcaacatag

19

<210> 71  
<211> 20  
<212> ADN  
<213> Artificial Sequence  
<220>  
<223> Amorce



<400> 71  
catcttcaag tttggtgcac

20

<210> 72  
<211> 1689  
<212> ADN  
<213> Mus musculus  
<220>  
<223> ADN Spg4 incomplet.

<400> 72  
aggccgagag cgtccgcgtc ttccacaagc aggccttcga gtacatctcc attgccctgc 60  
gcatcgacga ggaagagaaa gcaggacaga aggaacaagc tgtggaatgg tataagaaag 120  
gtatcgaaaga actggaaaaa ggaatcgctg ttatagttac gggccaaggt gaacagtatg 180  
aaagagctag acgtcttcaa gccaaaatga tgactaattt agttatggcc aaggaccgtt 240  
tacaacttct agagaagctg caaccagttt tgcaattttc caagtcacag acggacgtct 300  
ataacgagag tactaacctg acatgccgca atggacatct ccagtcagaa agtggagcag 360  
ttccgaagag gaaagacccc ttaacacatg ctagtaattc attgcctcga tcaaaaactg 420  
tcctgaaaag tggctccgca gggctctccg gtcaccacag ggcgcctagt tgcagtgggt 480  
tgtccatggt ttctggagca agaccgggac ctggtcctgc agctaccaca cataagggtta 540  
ctccaaaacc aaatagaacc aacaaacctt ctactcccac aactgcagtt cggaaaaaga 600  
aagacttgaa aaatttttagg aatgtggaca gcaatcttgc taaccttata atgaatgaaa 660  
ttgttgacaa tgggacagct gttaagtttg atgacatagc cgggcaggag ctggcaaagc 720  
aagcgtgca ggagattgtc atccttccct ctctgcggcc tgagttgttc acaggggtca 780  
gagctcctgc tagaggcttg ttactcttcg gtccgccagg aaacggaaaa acaatgctgg 840  
ctaaagcagt agctgcagag tctaattgca cctttttcaa cataagtgtc gccagtttaa 900  
cttcaaaata tgtgggagaa ggagagaaat tggtgagagc tctctttgct gtggctcgag 960  
aacttcaacc atctataatt tttatagatg aagttgacag tcttttgtgt gagagacggg 1020  
aaggggagca cgacgttagc agacggctaa agacggaatt ttaatagaa tttgacgggg 1080  
tgcaatctgc tggagatgac agagtacttg taatgggtgc aactaacagc cccaagagc 1140  
ttgatgaagc tgttctcagg cgtttcatta aacgggtata tgtgtcctta ccaaagagg 1200  
agacaagact cttctgctt aaaaacctgt tgtgtaaaca aggaagtcca ctgaccctaa 1260  
aagaactcgc acagcttgct agaattgacc atggatactc tgggaagtgt ctgaccgctt 1320  
tggccaagga tgcagccctg ggtcctatcc gagaactgaa gccagagcag gtgaagaata 1380  
tgtctgccag tgagatgaga aatattcgat tatctgactt cacagaatcc ttaaaaaaga 1440  
taaaacgcag tgtgagtcct cagaccttag aagcatatcat acgctggaac aaggattttg 1500  
gagacaccac tgtttaaagg aatggatgcc tctgtgagcc catagaacat cgcacttcac 1560  
aggaaacaag agctttggct acaggaaccc agacttcggt tacaggacgt tttagagttt 1620  
tcatttttgt gcaccaaact tgaagaggaa caagaagaca gacctaaata aaatatgcaa 1680  
tatgaatgg 1689

<210> 73  
<211> 504  
<212> PRT  
<213> Mus musculus  
<220>  
<223> Spastin murine incomplète.

<400> 73  
Ala Glu Ser Val Arg Val Phe His Lys Gln Ala Phe Glu Tyr Ile Ser  
1 5 10 15  
Ile Ala Leu Arg Ile Asp Glu Glu Glu Lys Ala Gly Gln Lys Glu Gln  
20 25 30  
Ala Val Glu Trp Tyr Lys Lys Gly Ile Glu Glu Leu Glu Lys Gly Ile  
35 40 45  
Ala Val Ile Val Thr Gly Gln Gly Glu Gln Tyr Glu Arg Ala Arg Arg  
50 55 60  
Leu Gln Ala Lys Met Met Thr Asn Leu Val Met Ala Lys Asp Arg Leu  
65 70 75 80



Gln	Leu	Leu	Glu	Lys	Leu	Gln	Pro	Val	Leu	Gln	Phe	Ser	Lys	Ser	Gln
Thr	Asp	Val	Tyr	Asn	Glu	Ser	Thr	Asn	Leu	Thr	Cys	Arg	Asn	Gly	His
Leu	Gln	Ser	Glu	Ser	Gly	Ala	Val	Pro	Lys	Arg	Lys	Asp	Pro	Leu	Thr
His	Ala	Ser	Asn	Ser	Leu	Pro	Arg	Ser	Lys	Thr	Val	Leu	Lys	Ser	Gly
Ser	Ala	Gly	Leu	Ser	Gly	His	His	Arg	Ala	Pro	Cys	Ser	Gly	Leu	Thr
Ser	Met	Val	Ser	Gly	Ala	Arg	Pro	Gly	Pro	Gly	Pro	Ala	Ala	Thr	Thr
His	Lys	Gly	Thr	Pro	Lys	Pro	Asn	Arg	Thr	Asn	Lys	Pro	Ser	Thr	Pro
Thr	Thr	Ala	Val	Arg	Lys	Lys	Lys	Asp	Leu	Lys	Asn	Phe	Arg	Asn	Val
Asp	Ser	Asn	Leu	Ala	Asn	Leu	Ile	Met	Asn	Glu	Ile	Val	Asp	Asn	Gly
Thr	Ala	Val	Lys	Phe	Asp	Asp	Ile	Ala	Gly	Gln	Glu	Leu	Ala	Lys	Gln
Ala	Leu	Gln	Glu	Ile	Val	Ile	Leu	Pro	Ser	Leu	Arg	Pro	Glu	Leu	Phe
Thr	Gly	Leu	Arg	Ala	Pro	Ala	Arg	Gly	Leu	Leu	Leu	Phe	Gly	Pro	Pro
Gly	Asn	Gly	Lys	Thr	Met	Leu	Ala	Lys	Ala	Val	Ala	Ala	Glu	Ser	Asn
Ala	Thr	Phe	Phe	Asn	Ile	Ser	Ala	Ala	Ser	Leu	Thr	Ser	Lys	Tyr	Val
Gly	Glu	Gly	Glu	Lys	Leu	Val	Arg	Ala	Leu	Phe	Ala	Val	Ala	Arg	Glu
Leu	Gln	Pro	Ser	Ile	Ile	Phe	Ile	Asp	Glu	Val	Asp	Ser	Leu	Leu	Cys
Glu	Arg	Arg	Glu	Gly	Glu	His	Asp	Ala	Ser	Arg	Arg	Leu	Lys	Thr	Glu
Phe	Leu	Ile	Glu	Phe	Asp	Gly	Val	Gln	Ser	Ala	Gly	Asp	Asp	Arg	Val
Leu	Val	Met	Gly	Ala	Thr	Asn	Arg	Pro	Gln	Glu	Leu	Asp	Glu	Ala	Val
Leu	Arg	Arg	Phe	Ile	Lys	Arg	Val	Tyr	Val	Ser	Leu	Pro	Asn	Glu	Glu
Thr	Arg	Leu	Leu	Leu	Leu	Lys	Asn	Leu	Leu	Cys	Lys	Gln	Gly	Ser	Pro
Leu	Thr	Gln	Lys	Glu	Leu	Ala	Gln	Leu	Ala	Arg	Met	Thr	Asp	Gly	Tyr
Ser	Gly	Ser	Asp	Leu	Thr	Ala	Leu	Ala	Lys	Asp	Ala	Ala	Leu	Gly	Pro
Ile	Arg	Glu	Leu	Lys	Pro	Glu	Gln	Val	Lys	Asn	Met	Ser	Ala	Ser	Glu
Met	Arg	Asn	Ile	Arg	Leu	Ser	Asp	Phe	Thr	Glu	Ser	Leu	Lys	Lys	Ile
Lys	Arg	Ser	Val	Ser	Pro	Gln	Thr	Leu	Glu	Ala	Tyr	Ile	Arg	Trp	Asn
Lys	Asp	Phe	Gly	Asp	Thr	Thr	Val								

```
<210> 74
<211> 24
<212> ADN
<213> Homo sapiens
```



&lt;220&gt;

&lt;223&gt; Site accepteur d'épissage du gène SPG4.

&lt;400&gt; 74

atttttttatt ttaaagcagg acag

24

&lt;210&gt; 75

&lt;211&gt; 24

&lt;212&gt; ADN

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;223&gt; Site accepteur d'épissage du gène SPG4.

&lt;400&gt; 75

aattttttttc tttcaggtga acag

24

&lt;210&gt; 76

&lt;211&gt; 24

&lt;212&gt; ADN

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;223&gt; Site accepteur d'épissage du gène SPG4.

&lt;400&gt; 76

cttctctgtt gcatagagaa gatg

24

&lt;210&gt; 77

&lt;211&gt; 24

&lt;212&gt; ADN

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;223&gt; Site accepteur d'épissage du gène SPG4.

&lt;400&gt; 77

actttttcct tgtcagaaag tgga

24

&lt;210&gt; 78

&lt;211&gt; 24

&lt;212&gt; ADN

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;223&gt; Site accepteur d'épissage du gène SPG4.

&lt;400&gt; 78

ttttgtatcc tttaagggtg ctcc

24

&lt;210&gt; 79

&lt;211&gt; 24

&lt;212&gt; ADN

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;223&gt; Site accepteur d'épissage du gène SPG4.

&lt;400&gt; 79

aggtcttggt tcttagtgga acag

24

&lt;210&gt; 80

&lt;211&gt; 24

&lt;212&gt; ADN



<213> Homo sapiens  
<220>  
<223> Site accepteur d'épissage du gène SPG4.

<400> 80  
agtatatatt ttttagttgt tcac 24

<210> 81  
<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site accepteur d'épissage du gène SPG4.

<400> 81  
cttgtgattt ttaaaggcta aagc 24

<210> 82  
<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site accepteur d'épissage du gène SPG4.

<400> 82  
taatgctttg ttttaggtgg gaga 24

<210> 83  
<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site accepteur d'épissage du gène SPG4.

<400> 83  
cttgtatttc ctctagatga agtt 24

<210> 84  
<211> 24  
<212> ADN  
<213> Homo sapiens  
  
<220>  
<223> Site accepteur d'épissage du gène SPG4.

<400> 84  
gattttttgc ttgtaggtac agtc 24

<210> 85  
<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site accepteur d'épissage du gène SPG4.

<400> 85  
ggattttttt ttttaggcgt ttca 24

<210> 86



<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site accepteur d'épissage du gène SPG4.

<400> 86  
ctcaatatt tttcagacaa gaçt 24

<210> 87  
<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site accepteur d'épissage du gène SPG4.

<400> 87  
cctt cctcagaatg actg 24

<210> 88  
<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site accepteur d'épissage du gène SPG4.

<400> 88  
cttttatggt ttacagaact aaaa 24

<210> 89  
<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site accepteur d'épissage du gène SPG4.

<400> 89  
ctttttaaaa atctagatga gaaa 24

<210> 90  
<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site donneur d'épissage du gène SPG4.

<400> 90  
tgagaaaggt aactaggggg ctgg 24

<210> 91  
<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site donneur d'épissage du gène SPG4.

<400> 91  
aggacaaggt aagattgtat ttgt 24



<210> 92  
<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site donneur d'épissage du gène SPG4.

<400> 92  
acttctaggt atcaattaat gtat 24

<210> 93  
<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site donneur d'épissage du gène SPG4.

<400> 93  
ccagtcaggt gggtttaggt taac 24

<210> 94  
<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site donneur d'épissage du gène SPG4.

<400> 94  
ctcataaggt attctgggac agta 24

<210> 95  
<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site donneur d'épissage du gène SPG4.

<400> 95  
gtggacaagt aagttttgcc atct 24

<210> 96  
<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site donneur d'épissage du gène SPG4.

<400> 96  
ggcctgaggt aagaacttta tatt 24

<210> 97  
<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site donneur d'épissage du gène SPG4.

<400> 97  
caatgctgggt aagggttctc ttca 24



<210> 98  
<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site donneur d'épissage du gène SPG4.

<400> 98  
caaaatacgt gagtgtctctg tttc 24

<210> 99  
<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site donneur d'épissage du gène SPG4.

<400> 99  
ttttataggt aagaacatat tttc 24

<210> 100  
<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site donneur d'épissage du gène SPG4.

<400> 100  
ttgatggtgt aagtgttgat tatg 24

<210> 101  
<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site donneur d'épissage du gène SPG4.

<400> 101  
gttctcaggt aggagattt atat 24

<210> 102  
<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site donneur d'épissage du gène SPG4.

<400> 102  
atgaggaggt atgtatctgt gttt 24

<210> 103  
<211> 24  
<212> ADN  
<213> Homo sapiens  
<220>  
<223> Site donneur d'épissage du gène SPG4.

<400> 103  
cttgctaggt gagtaatttg gatt 24



<210> 104  
 <211> 24  
 <212> ADN  
 <213> Homo sapiens  
 <220>  
 <223> Site donneur d'épissage du gène SPG4.

<400> 104  
 tatccgaggt aggtatacaa gagc

24

<210> 105  
 <211> 24  
 <212> ADN  
 <213> Homo sapiens  
 <220>  
 <223> Site donneur d'épissage du gène SPG4.

<400> 105  
 ccagtgaggt atagtatttt acaa

24

<210> 106  
 <211> 2152  
 <212> ADN  
 <213> Mus musculus  
 <220>  
 <223> ADN Spg4 complet

<400> 106  
 gctccccgggc tccggcgggc gcgcgagcgg ctccgtgggc ccccgccgcc gcagtggcag 60  
 tgcccgccgc cgccgcttgg tcgccgtcgg tctgcgggaa gggggttatg gcggcgccgg 120  
 cagtgggagc tgtgaatgag ttctccggcc ggacgacgga agaagaaagg ctccggcgcc 180  
 gcgagcccgg cgcccgccag gcctccgccc cccgcgcggg tccccgcccc tgccgcccgg 240  
 ccggccccctg cggccgggctc gccgcctaag cggaacctgt cttctttctc gtccccgctg 300  
 gtcgtcggct tcgccctgct gcgcctgctg gcctgccacc tggggctcct ctccgcgtgg 360  
 ctctgccagc gcttctcccc cgccctcatg gccgccaaga ggagctccgg gaccgcgccg 420  
 gcgccgcct cgccctcgcc ccagcgccc ggaccgggtg gcgaggccga gagcgtccgc 480  
 gtcttccaca agcaggcctt cgagtacatc tccattgccc tgcgcacga cgaggaagag 540  
 aaagcaggac agaaggaaca agctgtggaa tggataaga aaggtatcga agaactggaa 600  
 aaaggaatcg ctgttatagt tacgggccaa ggtgaacagt atgaaagagc tagacgtctt 660  
 caagccaaaa tgatgactaa tttagttagt gccaaaggacc gtttacaact tctagagaag 720  
 ctgcaaccag ttttgcaatt ttccaagtca cagacggacg tctataacga gactactaac 780  
 ctgacatgcc gcaatggaca tctccagtca gaaagtggag cagttccgaa gaggaagac 840  
 cccttaacac atgctagtaa ttcattgcct cgatcaaaaa ctgtcctgaa aagtggctcc 900  
 gcagggctct ccggtcacca cagggcgcct agttgcagtg gtttgtccat ggtttctgga 960  
 gcaagaccgg gacctggtcc tgcagctacc acacataagg gtactccaaa accaaataga 1020  
 accaacaac cttctactcc cacaactgca gttcgaaaa agaaagactt gaaaaatttt 1080  
 aggaatgtgg acagcaatct tgctaacctt ataatgaatg aaattgttga caatgggaca 1140  
 gctgttaagt ttgatgacat agccgggagc gagctggcaa agcaagcgt gcaggagatt 1200  
 gtcactcttc cttctctgcg gcctgagttg ttcacagggc tcagagctcc tgctagaggc 1260  
 ttgttactct tcggtccgcc aggaaacgga aaaacaatgc tggctaaagc agtagctgca 1320  
 gagtctaata cgaccttttt caacataagt gctgccagtt taacttcaaa atatgtggga 1380  
 gaaggagaga aattggtgag agctctcttt gctgtggctc gagaacttca accatctata 1440  
 atttttatag atgaagttga cagtcttttg tgtgagagac gggaaagggg gcacgacgt 1500  
 agcagagcgc taaagacgga atttttaata gaatttgacg gggtgcaatc tgctggagat 1560  
 gacagagtac ttgtaattgg tgcaactaac aggcccaag agcttgatga agctgttctc 1620  
 aggcgtttca ttaaaccggg atatgtgtcc aggagacaag actccttctg 1680  
 cttaaaaaacc tgttgtgtaa acaaggaagt ccactgaccc aaaaagaact cgcacagctt 1740  
 gctagaatga ccgatggata ctctggaagt gatctgaccg ctttggccaa ggatgcagcc 1800  
 ctgggtccta tccgagaact gaagccagag cagggtgaaga atatgtctgc cagtgagatg 1860



```

agaaatattc gattatctga cttcacagaa tccttaaaaa agataaaacg cagtgtgagt 1920
cctcagacct tagaagcata catacgctgg aacaaggatt ttggagacac cactgtttta 1980
aggaatggat gcctctgtga gcccatagaa catcgcaactt cacaggaaac aagagctttg 2040
gctacaggaa cccagacttc gtttacagga cgttttagag ttttcatttt tgtgcaccaa 2100
acttgaagag gaacaagaag acagacctaa ataaaatatg caatatgaat gg 2152

```

&lt;210&gt; 107

&lt;211&gt; 614

&lt;212&gt; PRT

&lt;213&gt; Mus musculus

&lt;220&gt;

&lt;223&gt; Spastin murine complète

&lt;400&gt; 107

```

Met Ser Ser Pro Ala Gly Arg Arg Lys Lys Lys Gly Ser Gly Gly Ala
 1          5          10          15
Ser Pro Ala Pro Ala Arg Pro Pro Pro Pro Ala Ala Val Pro Ala Pro
      20          25          30
Ala Ala Gly Pro Ala Pro Ala Ala Gly Ser Pro Pro Lys Arg Asn Leu
      35          40          45
Ser Ser Phe Ser Ser Pro Leu Val Val Gly Phe Ala Leu Leu Arg Leu
      50          55          60
Leu Ala Cys His Leu Gly Leu Leu Phe Ala Trp Leu Cys Gln Arg Phe
      65          70          75          80
Ser Arg Ala Leu Met Ala Ala Lys Arg Ser Ser Gly Thr Ala Pro Ala
      85          90          95
Pro Ala Ser Pro Ser Pro Pro Ala Pro Gly Pro Gly Gly Glu Ala Glu
      100          105          110
Ser Val Arg Val Phe His Lys Gln Ala Phe Glu Tyr Ile Ser Ile Ala
      115          120          125
Leu Arg Ile Asp Glu Glu Glu Lys Ala Gly Gln Lys Glu Gln Ala Val
      130          135          140
Glu Trp Tyr Lys Lys Gly Ile Glu Glu Leu Glu Lys Gly Ile Ala Val
      145          150          155          160
Ile Val Thr Gly Gln Gly Glu Gln Tyr Glu Arg Ala Arg Arg Leu Gln
      165          170          175
Ala Lys Met Met Thr Asn Leu Val Met Ala Lys Asp Arg Leu Gln Leu
      180          185          190
Leu Glu Lys Leu Gln Pro Val Leu Gln Phe Ser Lys Ser Gln Thr Asp
      195          200          205
Val Tyr Asn Glu Ser Thr Asn Leu Thr Cys Arg Asn Gly His Leu Gln
      210          215          220
Ser Glu Ser Gly Ala Val Pro Lys Arg Lys Asp Pro Leu Thr His Ala
      225          230          235          240
Ser Asn Ser Leu Pro Arg Ser Lys Thr Val Leu Lys Ser Gly Ser Ala
      245          250          255
Gly Leu Ser Gly His His Arg Ala Pro Ser Cys Ser Gly Leu Ser Met
      260          265          270
Val Ser Gly Ala Arg Pro Gly Pro Gly Pro Ala Ala Thr Thr His Lys
      275          280          285
Gly Thr Pro Lys Pro Asn Arg Thr Asn Lys Pro Ser Thr Pro Thr Thr
      290          295          300
Ala Val Arg Lys Lys Lys Asp Leu Lys Asn Phe Arg Asn Val Asp Ser
      305          310          315          320
Asn Leu Ala Asn Leu Ile Met Asn Glu Ile Val Asp Asn Gly Thr Ala
      325          330          335
Val Lys Phe Asp Asp Ile Ala Gly Gln Glu Leu Ala Lys Gln Ala Leu
      340          345          350

```







.

.

.

.

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/FR 00/02433

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/12 C12N5/10 C07K14/47 C07K16/18 C12Q1/68  
G01N33/50 A01K67/027

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K A01K G01N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EMBL, STRAND, BIOSIS

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>KIKUNO R. ET AL.: "Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 cDNA clones from brain which code for large proteins in vitro." DNA RES., vol. 6, 30 June 1999 (1999-06-30), pages 197-205, XP000852618 the whole document -&amp; DATABASE EMBL 'Online! Accession No. AB029006, 4 August 1999 (1999-08-04) OHARA O.: "Homo sapiens mRNA for KIAA1083 protein" XP002140994 the whole document</p> <p style="text-align: center;">--- -/--</p>	<p>1,4,5, 14,16, 18-21, 24-30</p>



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

## \* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*Z\* document member of the same patent family

Date of the actual completion of the international search

4 January 2001

Date of mailing of the international search report

25/01/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Gal11, I

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
x	<p>DATABASE EMBL SEQUENCES 'Online! Accession No. HS1179930, 29 March 1997 (1997-03-29) MYERS R.M.: "human STS SHGC-44567" XP002156510 brin complémentaire aux nucléotides 2204-2575 de la séquence no. 2</p> <p>HAZAN J. ET AL.: "Spastin, a new AAA protein, is altered in the most frequent form of autosomal dominant spastic paraplegia" NAT. GENET., vol. 23, November 1999 (1999-11), pages 296-303, XP000914979 the whole document</p> <p>HEINZLEF O. ET AL.: "Mapping of a complicated familial spastic paraplegia to locus SPG4 on chromosome 2p" J. ED. GENET., vol. 35, no. 2, February 1998 (1998-02), pages 89-93, XP000914971 cited in the application the whole document</p>	<p>1,4,5, 14,16, 18-21, 24-30</p> <p>1-36</p> <p>1-36</p>

## Continuation of Box I.2

Claims 13-15, 19-36 partly concern nucleic acid sequences promoting and/or regulating the expression of the SPG4 gene, as well as the corresponding polypeptides, without providing therein a real technical characterisation. Said claims concern also nucleic acids and polypeptides derived from any other gene controlling SPG4, which is neither defined nor characterised in the description. Consequently, said claims are ambiguous and vague, and their subject matter is not sufficiently disclosed and supported in accordance with PCT Article 5 and 6. Consequently, it not possible to carry out any meaningful search for speculative claims whereof the subject matter consists merely of a list of results to be obtained in future.

The applicant's attention is drawn to the fact that claims, or parts of claims, in respect of which no search report has been established need not be the subject of a preliminary examination report (PCT Rule 66.1(e)). The applicant is advised that the line of conduct adopted by the EPO acting in its capacity as International Preliminary Examination Authority is not to proceed with an examination of a subject matter in respect of which no search has been carried out. Such an attitude will remain unchanged, notwithstanding whether the claims have been modified or not, either after the report has been received or during any procedure under Chapter II.



5

•

•

•

## RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale No

PCT/FR 00/02433

## A. CLASSEMENT DE L'OBJET DE LA DEMANDE

CIB 7 C12N15/12 C12N5/10 C07K14/47 C07K16/18 C12Q1/68  
G01N33/50 A01K67/027

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

## B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)

CIB 7 C12N C07K A01K G01N C12Q

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si réalisable, termes de recherche utilisés)

EMBL, STRAND, BIOSIS

## C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie *	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	<p>KIKUNO R. ET AL.: "Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 cDNA clones from brain which code for large proteins in vitro." DNA RES., vol. 6, 30 juin 1999 (1999-06-30), pages 197-205, XP000852618 le document en entier -&amp; DATABASE EMBL 'en ligne! Accession No. AB029006, 4 août 1999 (1999-08-04) OHARA O.: "Homo sapiens mRNA for KIAA1083 protein" XP002140994 le document en entier ---</p>	<p>1,4,5, 14,16, 18-21, 24-30</p>

☒ Voir la suite du cadre C pour la fin de la liste des documents☐ Les documents de familles de brevets sont indiqués en annexe

## \* Catégories spéciales de documents cités:

- \*A\* document définissant l'état général de la technique, non considéré comme particulièrement pertinent
- \*E\* document antérieur, mais publié à la date de dépôt international ou après cette date
- \*L\* document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)
- \*O\* document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens
- \*P\* document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

- \*T\* document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention
- \*X\* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément
- \*Y\* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier
- \*Z\* document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

4 janvier 2001

Date d'expédition du présent rapport de recherche internationale

25/01/2001

Nom et adresse postale de l'administration chargée de la recherche internationale

Office Européen des Brevets, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Fonctionnaire autorisé

Galli, I

## C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	DATABASE EMBL SEQUENCES 'en ligne! Accession No. HS1179930, 29 mars 1997 (1997-03-29) MYERS R.M.: "human STS SHGC-44567" XP002156510 brin complémentaire aux nucléotides 2204-2575 de la séquence no. 2 ---	1,4,5, 14,16, 18-21, 24-30
P,X	HAZAN J. ET AL.: "Spastin, a new AAA protein, is altered in the most frequent form of autosomal dominant spastic paraplegia" NAT. GENET., vol. 23, novembre 1999 (1999-11), pages 296-303, XP000914979 le document en entier ---	1-36
A	HEINZLEF O. ET AL.: "Mapping of a complicated familial spastic paraplegia to locus SPG4 on chromosome 2p" J. MED. GENET., vol. 35, no. 2, février 1998 (1998-02), pages 89-93, XP000914971 cité dans la demande le document en entier -----	1-36

## SUITE DES RENSEIGNEMENTS INDIQUES SUR PCT/SA/ 210

## Suite du cadre I.2

Les revendications 13-15, 19-36 concernent en partie des séquences d'acides nucléiques promotrices et/ou régulatrices de l'expression du gène SPG4, ainsi que des polypeptides correspondants, sans y donner une véritable caractérisation technique. Ces revendications comprennent donc aussi des nucléiques et polypeptides dérivés de tout autre gène contrôlant SPG4, ce qui n'est ni défini ni caractérisé dans la description. En conséquence, ces revendications sont ambiguës et vagues, et leur sujet-matière n'est pas suffisamment décrit et démontré en accord avec les Articles 5 et 6 TCP. Aucune recherche ne peut être établie pour des revendications spéculatives dont l'objet ne serait qu'une simple liste de résultats à obtenir dans le futur.

L'attention du déposant est attirée sur le fait que les revendications, ou des parties de revendications, ayant trait aux inventions pour lesquelles aucun rapport de recherche n'a été établi ne peuvent faire obligatoirement l'objet d'un rapport préliminaire d'examen (Règle 66.1(e) PCT). Le déposant est averti que la ligne de conduite adoptée par l'OEB agissant en qualité d'administration chargée de l'examen préliminaire international est, normalement, de ne pas procéder à un examen préliminaire sur un sujet n'ayant pas fait l'objet d'une recherche. Cette attitude restera inchangée, indépendamment du fait que les revendications aient ou n'aient pas été modifiées, soit après la réception du rapport de recherche, soit pendant une quelconque procédure sous le Chapitre II.

